SEMI REPRODUCING KERNEL HILBERT SPACES AND MIXED PRECISION COMPUTATION

A thesis submitted in partial fulfilment of the requirements for the Degree of

Masters of Science in Mathematics

in the University of Canterbury

Ronald Garing

University of Canterbury

2018

Abstract

Positive definite and conditionally positive definite functions are widely used in interpolation and smoothing problems, particularly when the data is scattered. This thesis concerns such functions and also concerns the somewhat related topics of reproducing kernel Hilbert spaces, and semi reproducing kernel Hilbert spaces. The thesis presents various pieces of the relevant theory, sometimes with known established methods of proof, and sometimes with novel proofs.

Chapter one concerns the history of a specific class of such functions, namely the radial basis functions.

Chapter two concerns the general properties of positive definite functions, highlighting their use in interpolation problems and establishing the existence of the corresponding native spaces.

Chapter three concerns reproducing kernel Hilbert Spaces and shows their relation to positive definite functions. Afterwards, interpolation and smoothing, along with other approximation problems are discussed within the reproducing kernel Hilbert space setting.

Chapter four concerns examples of positive definite functions in the settings of $\mathbb{R}^d$ and on the spheres $\mathbb{S}^{d-1}$.

Chapter five concerns the basic theory of conditionally positive definite functions and their application in interpolation problems.

Chapter six concerns a variant to reproducing kernel Hilbert spaces, namely their semi Hilbert space variant and discusses interpolation and smoothing in this setting.

Chapter seven concerns interpolation via Guassian functions and the use of higher precision arithmetic to counter poor conditioning. Unfortunately, due to lack of time, this chapter goes no further.

Lastly the Appendices contain the prerequisite information concerning semi inner product spaces and convex functions.

Acknowledgments

I would like to thank my supervisor Professor Rick Beatson for his constant support and supervision throughout this project. This has been an enjoyable year of learning.

I would also like to thank my parents for their constant encouragement.

Contents

1 Introduction

Consider interpolation and approximation in $\mathbb{R}^d$, $d \in \mathbb{N}$, by functions of the form

$$s(x) = \sum_{i=1}^{n} \lambda_i \phi(\|x - x_i\|) + p(x)$$

where

(1) $\phi : [0, \infty) \to \mathbb{R}$,

(2) $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_n) \in \mathbb{R}^n$,

(3) $\| \cdot \|$ is the 2-norm of $\mathbb{R}^d$,

(4) $X = \{x_1, x_2, \ldots, x_n\}$ is a set of $n \in \mathbb{N}$ scattered points in $\mathbb{R}^d$,

(5) p is a low degree polynomial.

With sufficient constraints on ϕ, $\boldsymbol{\lambda}$ and p, this is a radial basis function (RBF) interpolation problem. In the setting of $\mathbb{R}^1$ it has been known for a very long time that one can interpolate such data with a polynomial spline. So for example one could choose

$$\begin{aligned} \phi(x) &= |x|, \\ p &\in \pi_0^1, \end{aligned}$$

and find a piecewise linear interpolant. Similarly, one could choose

$$\begin{aligned} \phi(x) &= |x|^3, \\ p &\in \pi_1^1, \end{aligned}$$

and find say the natural cubic spline interpolant. This interpolant also minimises $\int_{x_1}^{x_n} s''(t)^2 dt$ over all interpolants (assuming here that $x_1 \leq x_2 \leq \ldots, \leq x_n$). In $\mathbb{R}^1$, questions as above were addressed in the study of splines and H-splines (see for example the texts of deBoor [4] and Schumacker [14]). As a third example, one could choose

$$\phi(x) = e^{-\alpha x^2}, \ \alpha > 0,$$

with no polynomial term, giving a Gaussian interpolant for s. This interpolant has the natural corresponding energy $(2\pi)^{-d/2} \int_{\mathbb{R}^d} \widehat{s}(t)^2 / \widehat{\phi}(t) dt$ of which in minimises over all suitably smooth functions s subject to the interpolation constraints. The work of Duchon (1977) [8] showed that minimal energy interpolation in $\mathbb{R}^d$ could be connected to RBF interpolation problems. This field has since blossomed, for an excellent text on this subject, see Wendland [18].

The examples above all give rise to conditionally positive definite functions in the form $\Phi : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$, $\Phi(x, y) = \phi(\|x - y\|)$. In this thesis I will discuss the general theory of symmetric positive definite functions and symmetric conditionally positive definite functions, along with their natural corresponding energy functions and native spaces. The use of these functions in interpolation, smoothing and mixed interpolation and smoothing will also be discussed. Examples of (conditionally) positive definite functions will be given. The somewhat connected theory of reproducing kernel Hilbert spaces and semi reproducing kernel Hilbert spaces will also be discussed.

2 Positive Definite Functions

2.1 Introduction

Let Ω be a non-empty set. A function $P : \Omega \times \Omega \to \mathbb{R}$ is **positive definite** (**PD**) over Ω if for any finite subset $X = \{x_1, x_2, \dots, x_n\} \subseteq \Omega$ of some size $n \in \mathbb{N}$ and any $\boldsymbol{a} = (a_1, a_2, \dots, a_n) \in \mathbb{R}^n$,

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j P(x_i, x_j) \geq 0. \tag{1}$$

This is equivalent to the $n \times n$ matrix

$$\boldsymbol{P}_X := \begin{bmatrix} P(x_1, x_1) & P(x_1, x_2) & \dots & P(x_1, x_n) \\ P(x_2, x_1) & P(x_2, x_2) & \dots & P(x_2, x_n) \\ \vdots & \ddots & \ddots & \vdots \\ P(x_n, x_1) & P(x_n, x_2) & \dots & P(x_n, x_n) \end{bmatrix}$$

being non-negative definite as

$$\boldsymbol{a}^T \boldsymbol{P}_X \boldsymbol{a} = \sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j P(x_i, x_j) \geq 0. \tag{2}$$

The above matrix $\boldsymbol{P}_X$ is the **Gramian matrix** of P at X. Common choices for Ω are $\mathbb{R}^d$ ($d \in \mathbb{N}$ dimension real space), $\mathbb{S}^{d-1}$ (the unit sphere of $\mathbb{R}^d$), compact subsets of $\mathbb{R}^d$ and simply connected complex domains. Furthermore, the usual positive definite functions of interest are continuous and reflect certain symmetry properties of the space Ω they are over (see Cheney & Light [6], Wendland [18]). P is **strictly positive definite** (**SPD**) over Ω if for any finite subset $X = \{x_1, x_2, \dots, x_n\} \subseteq \Omega$ of size $n \in \mathbb{N}$ and any $\boldsymbol{a} = (a_1, a_2, \dots, a_n) \in \mathbb{R}^n$, the inequality in (1) is strict whenever $\boldsymbol{a} \neq \boldsymbol{0}$. From (2), this is equivalent to the Gramian matrix $\boldsymbol{P}_X$ being positive definite. There is in fact a complex variant of a positive definite function to which the work below can be extended. Only the case of real valued positive definite functions will be considered here.

Our concern will be positive definite functions that are symmetric in the sense that

$$P(x, y) = P(y, x), \ \forall x, y \in \Omega.$$

This is equivalent to the Gramian matrix $\boldsymbol{P}_X$ being a symmetric matrix for any choice of X. Recall that SPD does not refer to "Symmetric Positive Definite", but "Strictly Positive Definite". Let $\delta_x : \mathbb{R}^\Omega \to \mathbb{R}, \ \ x \in \Omega$, be the point evaluation functional for x, defined as follows

$$\delta_x(f) = f(x), \ \ \forall f \in \mathbb{R}^\Omega,$$

where $\mathbb{R}^\Omega$ is the set of functions from Ω to $\mathbb{R}$. Throughout this section, P will be used to refer to a symmetric positive definite function.

For any linear functional μ over $\mathbb{R}^\Omega$, $\mu^{(1)}P$ and $\mu^{(2)}P$ denote the functions such that

$$\mu^{(1)}P(x) = \mu P(\,\cdot\,, x),$$
$$\mu^{(2)}P(x) = \mu P(x,\,\cdot\,),$$

where $P(\,\cdot\,, x)$ and $P(x,\,\cdot\,)$ are treated as functions of $\,\cdot\,$. So for any $x, y \in \Omega$, $\delta_x^{(1)}\delta_y^{(2)}P = P(x, y)$. For any $X = \{x_1, x_2, \ldots, x_n\} \subseteq \Omega$ of size $n \in \mathbb{N}$ and $\boldsymbol{a} = (a_1, a_2, \ldots, a_n) \in \mathbb{R}^n$

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j P(x_i, x_j) = (\sum_{i=1}^{n} a_i \delta_{x_i})^{(1)}(\sum_{j=1}^{n} a_j \delta_{x_j})^{(2)}P.$$

Given this, let Ω^* be the set of all finite linear combinations of point evaluations over $\mathbb{R}^\Omega$ (it follows that Ω^* is a real vector space). Letting $\mu \in \Omega^*$

$$\mu^{(1)}\mu^{(2)}P \geq 0.$$

Recall that P is assumed to be symmetric, so it defines a semi inner product (see Appendix A for definition) $\langle\cdot,\cdot\rangle_P : \Omega^* \times \Omega^* \to \mathbb{R}$ over Ω^* such that

$$\langle \mu_1, \mu_2\rangle_P = \mu_1^{(1)}\mu_2^{(2)}P,$$

for any $\mu_1, \mu_2 \in \Omega^*$. This semi inner product will be studied in section 2.4.

Positive definite functions inherit properties from non-negative definite matrices. For any $n \times n$ non-negative definite matrix $\boldsymbol{A}$

$$(\boldsymbol{a}^T\boldsymbol{A}\boldsymbol{b})^2 \leq (\boldsymbol{a}^T\boldsymbol{A}\boldsymbol{a})(\boldsymbol{b}^T\boldsymbol{A}\boldsymbol{b}),\ \forall \boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^n.$$

This follows simply from the fact that, letting $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^n$,

$$\begin{bmatrix} s & t \end{bmatrix} \begin{bmatrix} \boldsymbol{a}^T\boldsymbol{A}\boldsymbol{a} & \boldsymbol{a}^T\boldsymbol{A}\boldsymbol{b} \\ \boldsymbol{b}^T\boldsymbol{A}\boldsymbol{a} & \boldsymbol{b}^T\boldsymbol{A}\boldsymbol{b} \end{bmatrix} \begin{bmatrix} s \\ t \end{bmatrix} = (s\boldsymbol{a} + t\boldsymbol{b})^T\boldsymbol{A}(s\boldsymbol{a} + t\boldsymbol{b}) \geq 0,\ \forall s, t \in \mathbb{R}.$$

So the above 2×2 block matrix is symmetric non-negative definite implying that its determinant is non-negative, giving

$$(\boldsymbol{a}^T\boldsymbol{A}\boldsymbol{a})(\boldsymbol{b}^T\boldsymbol{A}\boldsymbol{b}) - (\boldsymbol{a}^T\boldsymbol{A}\boldsymbol{b})^2 \geq 0.$$

The equivalent theorem for symmetric PD functions follows.

Theorem 2.1.1. *Let Ω be a non-empty set and let P be a symmetric PD function over Ω. Let $X = \{x_1, x_2, \ldots, x_n\}, Y = \{y_1, y_2, \ldots, y_m\} \subseteq \Omega$ of size $n, m \in \mathbb{N}$ respectively and $\boldsymbol{a} = (a_1, a_2, \ldots, a_n) \in \mathbb{R}^n$, $\boldsymbol{b} = (b_1, b_2, \ldots, b_m) \in \mathbb{R}^m$. Then*

$$|\sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_j P(x_i, y_j)|^2 \leq \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j P(x_i, x_j) \sum_{k=1}^{m}\sum_{l=1}^{m} b_k b_l P(y_k, y_l).$$

Proof. Let $Z = X \cup Y = \{z_1, z_2, \ldots, z_N\}$. As X and Y may not be disjoint, it isn't necessarily the case that $N = n+m$. We wish to find a $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_N), \boldsymbol{\gamma} = (\gamma_1, \gamma_2, \ldots, \gamma_N) \in \mathbb{R}^N$ such that

$$\begin{aligned}
\boldsymbol{\lambda}^T \boldsymbol{P}_Z \boldsymbol{\gamma} &= \sum_{i=1}^{n} \sum_{j=1}^{m} a_i b_j P(x_i, y_j), \\
\boldsymbol{\lambda}^T \boldsymbol{P}_Z \boldsymbol{\lambda} &= \sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j P(x_i, x_j), \\
\boldsymbol{\gamma}^T \boldsymbol{P}_Z \boldsymbol{\gamma} &= \sum_{i=1}^{m} \sum_{j=1}^{m} b_i b_j P(y_i, y_j).
\end{aligned}$$

As $\boldsymbol{P}_Z$ is symmetric non-negative definite, $(\boldsymbol{\lambda}^T \boldsymbol{P}_Z \boldsymbol{\gamma})^2 \leq (\boldsymbol{\lambda}^T \boldsymbol{P}_Z \boldsymbol{\lambda})(\boldsymbol{\gamma}^T \boldsymbol{P}_Z \boldsymbol{\gamma})$. The appropriate choices for $\boldsymbol{\lambda}$ and $\boldsymbol{\gamma}$ are:

$$\boldsymbol{\lambda}_i = \begin{cases} a_k, & \text{if } z_i = x_k, \text{ for some } k = 1, \ldots, n, \\ 0, & \text{otherwise,} \end{cases},$$

$$\boldsymbol{\gamma}_i = \begin{cases} b_k, & \text{if } z_i = y_k, \text{ for some } k = 1, \ldots, m, \\ 0, & \text{otherwise.} \end{cases}.$$

This completes the proof. □

There is an immediate corollary to this theorem.

Corollary 2.1.1.1. *Let Ω be a non-empty set and let P be a symmetric PD function over Ω. Let $X = \{x_1, x_2, \ldots, x_n\} \subseteq \Omega$ of size $n \in \mathbb{N}$ and $\boldsymbol{a} = (a_1, a_2, \ldots, a_n) \in \mathbb{R}^n$. If*

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j P(x_i, x_j) = 0,$$

then

$$\sum_{i=1}^{n} a_i P(x_i, x) = 0, \ \forall x \in \Omega.$$

Proof. Let $x \in \Omega$. By Theorem 2.1.1,

$$|\sum_{i=1}^{n} a_i P(x_i, x)|^2 \leq \sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j P(x_i, x_j) P(x, x) = 0.$$

Hence

$$\sum_{i=1}^{n} a_i P(x_i, x) = 0.$$

This holds for any $x \in \Omega$. □

Let A, B be matrices of the same dimensions, the Hadamard product of A and B, denoted by $A \odot B$, is the entrywise product of A and B. That is to say

$$(A \odot B)_{ij} = (A)_{ij}(B)_{ij}.$$

Another property of non-negative definite matrices is the following: for any two non-negative definite matrices, their Hadamard product is non-negative definite. Furthermore, the Hadamard product of any two positive definite matrices is positive definite. This remarkable fact leads to an equivalent theorem for positive definite functions.

Theorem 2.1.2. *Let Ω be a non-empty set and let P and Q be PD functions over Ω. The product of P and Q is PD over Ω. Furthermore, if both P and Q are SPD functions over Ω, then their product is SPD over Ω.*

Proof. Let $R : \Omega \times \Omega \to \mathbb{R}$ such that

$$R(x, y) = P(x, y)Q(x, y), \ \forall x, y \in \Omega.$$

We wish to show that R is PD over Ω. Let $X = \{x_1, x_2, \ldots, x_n\} \subseteq \Omega$ of size $n \in \mathbb{N}$. $\boldsymbol{P}_X \odot \boldsymbol{Q}_X =: \boldsymbol{R}_X$, where $\odot$ denotes the Hadamard product. So $\boldsymbol{R}_X$ is the Hadamard product of two non-negative definite matrices and as such is non-negative definite. This is equivalent to $R = PQ$ being PD over Ω. Furthermore, if P and Q are SPD over Ω, then the matrix $\boldsymbol{R}_X$ is the Hadamard product of two positive definite matrices and so $\boldsymbol{R}_X$ is positive definite. So R is SPD over Ω. □

As the non-negative scaling of a non-negative (positive) definite matrix is non-negative (positive) definite, and the sum of any two non-negative (positive) definite function is non-negative (positive) definite, the following theorem holds:

Theorem 2.1.3. *Let Ω be a non-empty set and let P and Q be (strictly) positive definite over Ω, the following hold:*

(1) sP is (strictly) positive definite, $\forall s \in \mathbb{R}_{>0}$.

(2) $P + Q$ is (strictly) positive definite.

It follows that the set of all positive definite functions over a non-empty set Ω is a convex cone that is closed under pointwise multiplication.

Note that the two theorems above did not suppose that the given positive definite functions where symmetric, yet Theorem 2.1.1 does. Recall that our main interest is symmetric PD functions.

Finally, let Ω_1 and Ω_2 be non-empty sets and let $\phi : \Omega_1 \to \Omega_2$ be an arbitrary function. For any PD function P over Ω_2 the function $(P \circ \phi)(x, y) = P(\phi(x), \phi(y))$ is trivially positive definite over Ω_1. Note however that if P is strictly positive definite, $P \circ \phi$ is not guaranteed to preserve the strict positive definiteness of P. Strictness is preserved however if ϕ is one-to-one.

2.2 Linear Interpolation with a Kernel

Let Ω be a non-empty set and $F : \Omega \times \Omega \to \mathbb{R}$. Suppose we wish to find a function $s : \Omega \to \mathbb{R}$ that satisfies the following interpolation constraints

$$s(x_i) = d_i, \ \forall i = 1, \dots, n,$$

where $X = \{x_1, x_2, \dots, x_n\} \subseteq \Omega$ of size $n \in \mathbb{N}$ and $\boldsymbol{d} = (d_1, d_2, \dots, d_n) \in \mathbb{R}^n$. Possibly such an s can be found in the form

$$s(\,\cdot\,) = \sum_{i=1}^{n} \lambda_i F(x_i,\,\cdot\,) \tag{3}$$

where $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \dots, \lambda_n) \in \mathbb{R}^n$. This would be equivalent to

$$\sum_{i=1}^{n} \lambda_i F(x_i, x_k) = d_k, \ \forall k = 1, \dots, n,$$

or expressed in matrix form

$$\begin{bmatrix} F(x_1,x_1) & F(x_2,x_1) & \dots & F(x_n,x_1) \\ F(x_1,x_2) & F(x_2,x_2) & \dots & F(x_n,x_2) \\ \vdots & \ddots & \ddots & \vdots \\ F(x_1,x_n) & F(x_1,x_n) & \dots & F(x_n,x_n) \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_n \end{bmatrix}.$$

So if the above $n \times n$ matrix is invertible, then a solution of the form (3) exists and is unique. If F where to be a symmetric SPD function over Ω then the above matrix is symmetric positive definite and so indeed it would be invertible. Additionally, one can then perform a Cholesky decomposition on the above matrix to solve the above equation reasonably efficiently. This motivates an interest in symmetric SPD functions. We will see however in chapter three that it is not just interpolation problems that SPD functions can be used for, but more general approximation problems.

2.3 Simple Construction of Positive Definite Functions

Let's start with a trivial construction of a class of PD functions. The method of proving that these functions are PD has analogues to other proofs of positive definiteness. Recall that SPD does not refer to "Symmetric Positive Definite", but "Strictly Positive Definite".

Theorem 2.3.1. *Let Ω be non-empty set and $\psi : \Omega \to \mathbb{R}$. The function P : $\Omega \times \Omega \to \mathbb{R}$ defined by*

$$P(x,y) = \psi(x)\psi(y), \ \forall x, y \in \Omega,$$

*is a symmetric PD function over Ω. Furthermore, P is **not** SPD over Ω if $|\Omega| > 1$.*

Proof. Symmetry follows trivially from the commutativity of multiplication. Let $X = \{x_1, x_2, \dots, x_n\} \subseteq \Omega$ of size $n \in \mathbb{N}$ and $\boldsymbol{a} = (a_1, a_2, \dots, a_n) \in \mathbb{R}^n$.

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j P(x_i, x_j) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \psi(x_i)\psi(x_j) = (\sum_{i=1}^{n} a_i \psi(x_i))^2 \geq 0.$$

So P is a symmetric PD function over Ω. Finally, consider the vector $\boldsymbol{v} = (\psi(x_1), \psi(x_2), \dots, \psi(x_n)) \in \mathbb{R}^n$. Letting $\boldsymbol{a}$ be a non-zero vector orthogonal to $\boldsymbol{v}$, which is guaranteed to exists if $n > 1$, we have

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j P(x_i, x_j) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \psi(x_i)\psi(x_j) = (\sum_{i=1}^{n} a_i \psi(x_i))^2 = (\boldsymbol{a}^T \boldsymbol{v})^2 = 0.$$

Hence P is not SPD over Ω when $|\Omega| > 1$. $\square$

In the case of a PD function over a set Ω of size one, any function of the form $F : \Omega \times \Omega \to \mathbb{R}$ is a constant function and so is trivially symmetric SPD so long as its range is not $\{0\}$.The above theorem is not very useful, especially for the construction of SPD functions, but it does demonstrate that PD functions are not necessarily exotic.

We will now move on to general methods for constructing PD functions. Note that none of these general methods here will be useful for finding interesting PD functions, that task will involve a study of the specific setting Ω. The methods below simply help in understanding PD functions. To the best of my knowledge these examples do not appear elsewhere.

Let Ω be a non-empty set of size $N \in \mathbb{N}$. Let's first show how to construct a PD function over Ω when Ω is a finite set. A simple way to construct a PD function P over Ω is given by the following method:

Method 1

(1) Give an arbitrary ordering $\{x_1, x_2, \dots, x_N\}$ of the elements in Ω.

(2) Choose an $N \times N$ non-negative definite matrix $\boldsymbol{A}$.

(3) Define $P : \Omega \times \Omega \to \mathbb{R}$ as follows:

$$P(x_i, x_j) = \boldsymbol{A}_{ij}, \;\; 1 \leq i, j \leq N.$$

Theorem 2.3.2. *Let $P : \Omega \times \Omega \to \mathbb{R}$ be defined via method 1. P is PD over Ω.*

Proof. Let $X \subseteq \Omega$ of size $n \in \mathbb{N}$. From step 1, every element of Ω has a corresponding index.

Let $\boldsymbol{A}$ be the non-negative definite matrix from step 2. For every $x \in X$, let $a_x \in \mathbb{R}$. For every element $y \in \Omega - X$, let $a_y = 0$. We will construct a vector $\boldsymbol{a} \in \mathbb{R}^N$ as follows. $\boldsymbol{a}_i = a_z$, where $z \in \Omega$ such that the index of z is i, $\forall i = 1, \ldots, N$. So we have that

$$\sum_{x \in X} \sum_{y \in X} a_x a_y P(x, y) = \sum_{i=1}^{N} \sum_{j=1}^{N} a_i a_j \boldsymbol{A}_{ij} = \boldsymbol{a}^T \boldsymbol{A} \boldsymbol{a} \geq 0.$$

Hence P is PD over Ω. $\square$

Noting the relationship between $N \times N$ symmetric, non-negative definite matrices and semi inner products over $\mathbb{R}^N$, an equivalent construction of a symmetric PD function P over Ω is:

Method 2

(1) Give an arbitrary ordering $\{x_1, x_2, \ldots, x_N\}$ of the elements in Ω.

(2) Choose a subset $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_N\}$ of $\mathbb{R}^N$.

(3) Choose an semi inner product $\langle \cdot, \cdot \rangle$ over $\mathbb{R}^N$.

(4) Define $P : \Omega \times \Omega \to \mathbb{R}$ as follows:

$$P(x_i, x_j) = \langle \boldsymbol{v}_i, \boldsymbol{v}_j \rangle, \;\; 1 \leq i, j \leq N.$$

It follows from method two that for any $X = \{x_1, x_2, \ldots, x_n\} \subseteq \Omega$ and any $\boldsymbol{a} = (a_1, a_2, \ldots, a_n) \in \mathbb{R}^n$ that

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j P(x_i, x_j) = \langle \sum_{i=1}^{n} a_i \boldsymbol{v}_i, \sum_{j=1}^{n} a_j \boldsymbol{v}_j \rangle \geq 0.$$

Now let's generalise this method to the setting where Ω isn't necessarily finite in size. Let Ω be a non-empty set of any size. We have the following method for constructing a symmetric PD function P over Ω:

Method 3

(1) Find a real semi inner product space $(V, \langle \cdot, \cdot \rangle)$.

(2) Choose a $\phi : \Omega \to V$.

(3) Define $P : X \times X \to \mathbb{R}$ as follows:

$$P(x, y) = \langle \phi(x), \phi(y) \rangle.$$

In the case where the real semi inner product space $(V, \langle \cdot, \cdot \rangle)$ from step 1 is a real Hilbert space, the function ϕ from step 2 is called a (real) **feature map** and $(V, \langle \cdot, \cdot \rangle)$ is its corresponding **feature space** (for a full treatment to feature maps and their corresponding spaces, see Schölkopf & Smola [16]).

Theorem 2.3.3. *Let Ω be a non-empty set. Let $P : \Omega \times \Omega \to \mathbb{R}$ be defined via method 3. P is symmetric PD over Ω.*

Proof. The symmetry of P follows from the symmetry of the semi inner product. Let $X = \{x_1, x_2, \ldots, x_n\} \subseteq \Omega$ of size $n \in \mathbb{N}$, $\boldsymbol{a} = (a_1, a_2, \ldots, a_n) \in \mathbb{R}^n$. Letting $(V, \langle \cdot, \cdot \rangle)$ and $\phi : \Omega \to V$ be from steps 1 and 2 respectively,

$$\begin{aligned}\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j P(x_i, x_j) &= \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \langle \phi(x_i), \phi(x_j) \rangle \\ &= \langle \sum_{i=1}^{n} a_i \phi(x_i), \sum_{i=1}^{n} a_j \phi(x_j) \rangle \geq 0.\end{aligned}$$

Hence, P is symmetric PD over Ω. ◻

It is natural to ask whether or not any symmetric PD functions can be constructed via method 3. The answer is not only yes, but for any symmetric PD function P over an arbitrary, non-empty Ω, there exists a feature space $(H_P, \langle \cdot, \cdot \rangle)$ and corresponding feature map $\phi : \Omega \to H_P$ such that $P(x, y) = \langle \phi(x), \phi(y) \rangle$. Proving this theorem will be the main result of this chapter. Before this, let's establish some properties of symmetric PD functions.

2.4 Properties of Symmetric Positive Definite Functions

It should be clear from the previous sections that symmetric PD functions have an intimate relationship with inner products. This fact will be explored here. From now on, all positive definite functions considered will be symmetric.

Let Ω be a non-empty set and let $P : \Omega \times \Omega \to \mathbb{R}$ be a symmetric PD function over Ω. For every element $x \in \Omega$, there is the function $P_x : \Omega \to \mathbb{R}$, $P_x(\,\cdot\,) = P(x,\,\cdot\,)$. The set

$$\mathcal{P} := \operatorname{span}\{P_x \mid x \in \Omega\}.$$

is the **space** of P.

Theorem 2.4.1. *Let Ω be a non-empty set and let P be a symmetric PD function over Ω. P is SPD over Ω if and only if the set $\{P_x \mid x \in \Omega\}$ is a Hamel basis of the space of P.*

Proof. Suppose P is SPD over Ω and let $X = \{x_1, x_2, \ldots, x_n\} \subseteq \Omega$ of size $n \in \mathbb{N}$, $\boldsymbol{a} = (a_1, a_2, \ldots, a_n) \in \mathbb{R}^n$ such that

$$\sum_{i=1}^{n} a_i P_{x_i} = 0.$$

It follows immediately that

$$\sum_{i=1}^{n}\sum_{i=1}^{n} a_i a_j P(x_i, x_j) = \sum_{j=1}^{n} a_j (\sum_{i=1}^{n} a_i P_{x_i}(x_j)) = \sum_{j=1}^{n} a_j \cdot 0 = 0.$$

As P is SPD, $\boldsymbol{a} = \boldsymbol{0}$. Thus $\{P_x \mid x \in \Omega\}$ is a Hamel basis for the space of P.

Now suppose $\{P_x \mid x \in \Omega\}$ is a Hamel basis for the space of P and let $X = \{x_1, x_2, \ldots, x_n\} \subseteq \Omega$ of size $n \in \mathbb{N}$, $\boldsymbol{a} = (a_1, a_2, \ldots, a_n) \in \mathbb{R}^n$ such that

$$\sum_{i=1}^{n} \sum_{i=1}^{n} a_i a_j P(x_i, x_j) = 0.$$

By Corollary 2.1.1.1,

$$\sum_{i=1}^{n} a_i P(x_i, x) = \sum_{i=1}^{n} a_i P_{x_i}(x) = 0, \ \forall x \in \Omega.$$

As $\{P_x \mid x \in \Omega\}$ is a Hamel basis for the space of P, $\boldsymbol{a} = \boldsymbol{0}$. Thus P is SPD. $\square$

Recall Ω^*, the set of all finite linear combinations of point evaluation functionals over $\mathbb{R}^\Omega$. It should be clear that every element of the space of a given symmetric PD function P can be expressed in the form $\mu^{(1)}P$, for some $\mu \in \Omega^*$. As P is symmetric, $\mu^{(1)}P = \mu^{(2)}P$. So, without ambiguity, we can refer to $\mu^{(1)}P$ as just μP.

Now let's establish the following lemma relating symmetric PD functions to inner products.

Lemma 2.4.2. *Let Ω be a non-empty set and let P be a symmetric PD function over Ω. The following hold:*

(1) $P(x, y)^2 \leq P(x, x)P(y, y), \ \forall x, y \in \Omega.$

(2) $P(x, x) \geq 0,$ *with equality when* $P_x = 0, \ \forall x \in \Omega.$

(3) $P(x, x) - 2P(x, y) + P(y, y) \geq 0,$ *with equality when* $P_x = P_y,$ $\forall x, y \in \Omega.$

Proof. *(1)* follows from Theorem 2.1.1. Alternatively, let $x, y \in \Omega$. As P is PD, the following matrix

$$\begin{bmatrix} P(x, x) & P(x, y) \\ P(y, x) & P(y, y) \end{bmatrix}$$

is non-negative definite and so its determinant is non-negative. This immediately gives us

$$P(x, x)P(y, y) - P(x, y)P(y, x) \geq 0.$$

Rearranging and simplifying gives *(1)*. From the definition of PD functions we see that

$$a_x^2 P(x, x) + 2a_x a_y P(x, y) + a_y^2 P(y, y) \geq 0,$$

where $a_x, a_y \in \mathbb{R} \backslash \{\boldsymbol{0}\}$. Setting $a_x = 1, a_y = 0$ gives

$$P(x, x) \geq 0.$$

Suppose the above holds for equality, then from *(1)*

$$P(x,y)^2 \leq P(x,x)P(y,y) = 0,$$

and so $P_x(y) = 0$. As y is arbitrary, $P_x = 0$, giving *(2)*. Setting $a_x = 1, a_y = -1$ gives

$$P(x,x) - 2P(x,y) + P(y,y) \geq 0,$$

Suppose the above holds for equality, then from Theorem 2.1.1

$$(P_x - P_y)(z)^2 \leq (P(x,x) - 2P(x,y) + P(y,y))(P(z,z)) = 0,$$

for any $z \in \Omega$. So $P_x = P_y$, giving *(3)*. $\square$

For the above theorem, *(1)* is analogous to the Cauchy-Schwarz inequality for inner products, *(2)* is analogous to the positive definiteness of inner products and *(3)* is analogous to the following : $\|v-u\|^2 = \|\boldsymbol{v}\|^2 - 2\langle \boldsymbol{v}, \boldsymbol{u}\rangle + \|\boldsymbol{u}\|^2 \geq 0$, with equality when $\boldsymbol{v} = \boldsymbol{u}$. Now let's completely establish the relationship between positive definite functions and inner products.

Recall again Ω^*, the set of all finite linear combinations of point evaluation functionals over $\mathbb{R}^\Omega$ and the semi inner product $\langle \cdot, \cdot \rangle_P$ defined over Ω^* as follows

$$\langle \mu_1, \mu_2 \rangle_P = \mu_1^{(1)} \mu_2^{(2)} P, \ \forall \mu_1, \mu_2 \in \Omega^*$$

Every element of the space $\mathcal{P}$ of P is expressible in the form $\mu P := \mu^{(1)} P$ where $\mu \in \Omega^*$. Consider the map $\phi : \Omega \to \mathcal{P}$, $\phi(x) = P_x$. This map can be viewed as a feature map by first defining the following inner product $\langle \cdot, \cdot \rangle : \mathcal{P} \times \mathcal{P} \to \mathbb{R}$:

$$\langle \mu_1 P, \mu_2 P \rangle = \langle \mu_1, \mu_2 \rangle_P = \mu_1^{(1)} \mu_2^{(2)} P.$$

Call this inner product the **natural** inner product of $\mathcal{P}$. This inner product is clearly bilinear and symmetric, however its positive definiteness is less obvious. This inner product is positive definite as $\langle \mu P, \mu P \rangle = \mu^{(1)} \mu^{(2)} P = 0$ implies $\mu P = 0$ by Corollary 2.1.1.1 for any $\mu \in \Omega^*$. We have the following important theorem concerning $(\mathcal{P}, \langle \cdot, \cdot \rangle)$.

Theorem 2.4.3. *Let Ω be a non-empty set and let P be a symmetric PD function over Ω. Let $(\mathcal{P}, \langle \cdot, \cdot \rangle)$ be the space of P equipped with its natural inner product. Point evaluation in $\mathcal{P}$ is continuous.*

Proof. Let $x \in \Omega$ and recall $\delta_x : \mathbb{R}^\Omega \to \mathbb{R}$, the point evaluation functional for x, defined as follows

$$\delta_x(f) = f(x), \ \forall f \in \mathcal{P}.$$

Then, letting $\mu P \in \mathcal{P}$, $\mu \in \Omega^*$,

$$|\delta_x(\mu P)| = |\mu P(x)| = |\langle \mu, \delta_x \rangle_P| = |\langle \mu P, P_x \rangle| \leq \|\mu P\| \|P_x\|,$$

where $\| \cdot \|$ is the norm of $\langle \cdot, \cdot \rangle$. Thus δ_x is bounded and so is continuous. $\square$

What follows is part the remarkable Moore-Aronszajin Theorem. Consider the completion of $(\mathcal{P}, \langle\cdot,\cdot\rangle)$ (say $(H_P, \langle\cdot,\cdot\rangle)$). As point evaluation functionals are continuous is $(\mathcal{P}, \langle\cdot,\cdot\rangle)$, they have unique continuous extensions in $(H_P, \langle\cdot,\cdot\rangle)$. So any element $f \in H$ can be seen as a function by defining $f(x) := \overline{\delta}_x f$ where $\overline{\delta}_x$ is the unique extension of $\delta_x \in \mathcal{P}^*$. H_P is the **Native Space** of P.

3 Reproducing Kernel Hilbert spaces

3.1 Introduction

An early comprehensive treatment of Reproducing Kernel Hilbert spaces is that of N. Aronszajn's in [1]. Let H be a real Hilbert space of functions from some non-empty set Ω to $\mathbb{R}$ with corresponding inner product $\langle\cdot,\cdot\rangle$ and norm $\|\cdot\|$. H is a **Reproducing Kernel Hilbert Space** (**RHKS**) over Ω if there is exists a function $K : \Omega \times \Omega \to \mathbb{R}$ such that

(1) $K(x,\ \cdot\ ) \in H,\ \ \forall x \in \Omega.$

(2) $\langle f, K(x,\ \cdot\ )\rangle = f(x),\ \ \forall f \in H,\ \ \forall x \in \Omega.$

K is called the **Reproducing Kernel** of H. For convenience, let K_x stand for $K(x,\ \cdot\ )$. The following then holds.

(1) $K_x \in H,\ \ \forall x \in \Omega.$

(2) $\langle f, K_x\rangle = f(x),\ \ \forall f \in H,\ \ \forall x \in \Omega.$

(3) $K(x,y) = \langle K_x, K_y\rangle,\ \ \forall x,y \in \Omega.$

(3) follows immediately from (1) and (2) and so is added in the list. It will be proved that the reproducing kernel of a RKHS is unique. In a RKHS, evaluating a function (say f) in H at some point (say x) in Ω is equivalent to taking the inner product of f with a unique function (say K_x) in H where K_x does not depending on f. Let $\delta_x : H \to \mathbb{R},\ \ x \in \Omega$, be the point evaluation functional for x, defined as follows

$$\delta_x(f) = f(x),\ \ \forall f \in H.$$

These functionals are clearly linear. If H is a RKHS with reproducing kernel K, then

$$|\delta_x(f)| = |f(x)| = |\langle f, K_x\rangle| \leq \|f\|\|K_x\|,\ \ \forall f \in H,$$

by Cauchy-Schwarz. Hence point evaluation is bounded in a RKHS and so $\delta_x \in H^*$, H^* being the set of all continuous linear functionals over H. In fact, the converse holds, that is to say

Theorem 3.1.1. *Let Ω be a non-empty set and let H be a real Hilbert space of functions over Ω. If point evaluation is continuous in H (that is if $\delta_x \in H^*,\ \ \forall x \in \Omega$), then H is a RKHS over Ω.*

Proof. By the Riesz representation theorem, for any $x \in \Omega$, as $\delta_x \in H^*$, there exists a function (say k_x) in H such that

$$\delta_x(f) = \langle f, k_x\rangle,\ \ \forall f \in H.$$

Define a function $K : \Omega \times \Omega \to \mathbb{R}$ by

$$K(x,\ \cdot\ ) = k_x.$$

Then clearly the following hold

(1) $K(x, \cdot) = k_x \in H, \;\; \forall x \in \Omega$.

(2) $\langle f, K(x, \cdot)\rangle = \langle f, k_x\rangle = \delta_x(f) = f(x), \;\; \forall f \in H, \;\; \forall x \in \Omega$.

Hence, K is a reproducing kernel for H. H is a RKHS. $\square$

The following equivalent characterisation of a RKHS has been established above.

Theorem 3.1.2. *Let Ω be a non-empty set and let H be a real Hilbert space of functions over Ω. H is a RKHS over Ω if and only if point evaluation is continuous in H.*

From now on, H will be a real RKHS with reproducing kernel K. Let $R : H^* \to H$ be the representation map of H. $R : H^* \to H$ is the reproducing mapping of H if

$$\langle f, R(\mu)\rangle = \mu f, \; \forall f \in H, \mu \in H^*.$$

There is a clear relationship between K and R, namely that $K_x = R(\delta_x)$, $\forall x \in \Omega$. K can then be represented as follows

$$K(x, y) = \langle K_x, K_y\rangle = \langle R(\delta_x), R(\delta_y)\rangle = R(\delta_x)(y) = (R \circ \delta)(x)(y).$$

So the reproducing kernel of H can be recognised as $R \circ \delta$.

Two basic properties of H and K will be established.

Theorem 3.1.3. *Let Ω be a non-empty set and let H be a RKHS over Ω with reproducing kernel K. The following hold for H and K:*

(1) H has a unique reproducing kernel.

(2) K is symmetric ($K(x, y) = K(y, x), \;\; \forall x, y \in \Omega$).

Proof. The uniqueness of the reproducing kernel of H follows from the uniqueness of representations of continuous linear functions via the Riesz representation theorem (that is to say that R is unique). Alternatively, Let K, K' be reproducing kernels for H and let $x \in \Omega$.

$$\begin{aligned} \|K_x - K'_x\|^2 &= \langle K_x - K'_x, K_x - K'_x\rangle \\ &= \langle K_x, K_x\rangle - \langle K_x, K'_x\rangle + \langle K'_x, K'_x\rangle - \langle K'_x, K_x\rangle \\ &= K_x(x) - K_x(x) + K'_x(x) - K'_x(x) = 0. \end{aligned}$$

So $K_x = K'_x$, $\forall x \in \Omega$, that is to say, $K = K'$.

The symmetry of K follows from the symmetry of the inner product. Let $y \in \Omega$.

$$K(x, y) = K_x(y) = \langle K_x, K_y\rangle = \langle K_y, K_x\rangle = K_y(x) = K(y, x).$$

So K is symmetric. $\square$

Reproducing kernels can be characterised independently of their corresponding RKHS. In fact, the reproducing kernels are precisely the symmetric positive definite functions (For definition of positive definite functions, see Chapter 2). This relationship between reproducing kernels and symmetric positive definite functions will be established by the following two theorems.

Theorem 3.1.4. *Let Ω be a non-empty set and let $(H, \langle\cdot,\cdot\rangle)$ be a RKHS over Ω with reproducing kernel K. K is symmetric positive definite over Ω. Furthermore, if the set $\{K_x \in H \mid x \in \Omega\}$ is linearly independent in H, then K is symmetric strictly positive definite.*

Proof. Symmetry follows from the previous theorem. Let $\{x_1, x_2, \ldots, x_n\} \subseteq \Omega$ be a set of $n \in \mathbb{N}$ distinct points of Ω. Let $\{a_1, a_2, \ldots, a_n\} \subset \mathbb{R}$. Consider the function

$$\sum_{i=1}^{n} a_i K_{x_i} \in H.$$

As $K_{x_i} \in H, \quad \forall i = 1, \ldots, n$, the above function is in H. Consider its' norm squared

$$\begin{aligned} 0 \leq \|\sum_{i=1}^{n} a_i K_{x_i}\|^2 = \langle\sum_{i=1}^{n} a_i K_{x_i}, \sum_{j=1}^{n} a_j K_{x_j}\rangle &= \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \langle K_{x_i}, K_{x_j}\rangle \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j K(x_i, x_j) \end{aligned}$$

So K is positive definite. For strictness, suppose the set $\{K_x \in H \mid x \in \Omega\}$ is linearly independent in H. Furthermore, suppose that

$$\|\sum_{i=1}^{n} a_i K_{x_i}\|^2 = 0.$$

This is equivalent to

$$\sum_{i=1}^{n} a_i K_{x_i} = 0, \text{ the zero function in } H.$$

As the set $\{K_x \in H \mid x \in \Omega\}$ is linearly independent in H, $a_i = 0, \quad \forall i = 1, \ldots, n$. Therefore, K is strictly positive definite. □

A sort of converse of this theorem holds. This is the Moore-Aronszajin Theorem.

Theorem 3.1.5. *Let Ω be a non-empty set. Let $P : \Omega \times \Omega \to \mathbb{R}$ be a symmetric positive definite function over Ω. There exists a unique RKHS H such that P is the reproducing kernel of H. Furthermore, if P is symmetric strictly positive definite, the set $\{P(x, \cdot) \mid x \in \Omega\}$ is linearly independent in H.*

Proof. The existence of a RKHS has been established at the end of Chapter 2. The native space $(H_P, \langle\cdot,\cdot\rangle)$ of P is a RHKS with reproducing kernel P.

For uniqueness, Let $(H, (\cdot,\cdot))$ be a RKHS with reproducing kernel P. The space of P, $\mathcal{P}$ (see section 2.4), must be a subset of H as $P_x \in H$ for any $x \in \Omega$ and so the linear combination of such functions must be in H, yet the set of linear combinations of such functions is $\mathcal{P}$. The inner product of H when restricted down to $\mathcal{P}$ must be the natural inner product of $\mathcal{P}$. so the completion of $\mathcal{P}$ (the native space of P) must be in H as H is complete. Let $f \in H$, $f = f_1 + f_2$ where $f_2 \in H_P$ and $(f_1, f_2) = 0$. For any $x \in \Omega$ we have

$$f(x) = (f, P_x) = (f_1 + f_2, P_x) = (f_2, P_x) = f_2(x).$$

So $f = f_2$ and $f \in H_P$. Thus $H = H_P$. Finally, Theorem 2.4.1 establishes that if P is SPD, the set $\{P(x,\ \cdot\ ) \mid x \in \Omega\}$ is linearly independent in $\mathcal{P}$, the space of P, and so in H_P. $\square$

3.2 Interpolation and Smoothing in Reproducing Kernel Hilbert Spaces

3.2.1 Interpolation with point evaluations

As shown before in section 2.2, one can interpolate with positive definite functions. For convenience, say that $(\Omega, H, \langle\cdot,\cdot\rangle, K)$ is a RKHS if Ω is an non-empty set and $(H, \langle\cdot,\cdot\rangle)$ is a RKHS over Ω with reproducing kernel K. $(\Omega, H, \langle\cdot,\cdot\rangle, K)$ is called **strict** if K is a **strictly** positive definite kernel. Recall from the previous subsection that K being strictly positive definite is equivalent to saying that the point evaluation functionals are all linearly independent in H^*. Interpolation can be seen as a norm minimisation problem.

Problem 1. *Let $(\Omega, H, \langle\cdot,\cdot\rangle, K)$ be a strict RKHS. Let $X = \{x_1, x_2, \ldots, x_n\} \subseteq \Omega$. Find $s \in H$ such that*

$$s = \operatorname*{arg\,min}_{f\in H} \|f\|^2 \text{ subject to } f(x_i) = d_i, \ \ \forall i = 1 \ldots n, \tag{4}$$

where $\boldsymbol{d} = (d_1, d_2, \ldots, d_n) \in \mathbb{R}^n$.

This problem has a unique solution. Before showing this, let's define and then prove the following. Let $(\Omega, H, \langle\cdot,\cdot\rangle, K)$ be a strict RKHS and let $X = \{x_1, x_2, \ldots, x_n\} \subseteq \Omega$ be $n \in \mathbb{N}$ distinct points in Ω. The $n \times n$ matrix $\boldsymbol{K}_X$ defined as follows

$$(\boldsymbol{K}_X)_{ij} = K(x_i, x_j)$$

is the **Gramian matrix** of K at X.

Lemma 3.2.1. *Let $(\Omega, H, \langle\cdot,\cdot\rangle, K)$ be a strict RKHS and let $X = \{x_1, x_2, \ldots, x_n\} \subseteq \Omega$ be $n \in \mathbb{N}$ distinct points in Ω. $\boldsymbol{K}_X$ is symmetric positive definite.*

Proof. This follows from K being symmetric strictly positive definite. Alternatively, let $\boldsymbol{a} = (a_1, a_2, \ldots, a_n) \in \mathbb{R}^n \backslash \{\boldsymbol{0}\}$.

$$\boldsymbol{a}^{\mathrm{T}} \boldsymbol{K}_X \boldsymbol{a} = \sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j K(x_i, x_j) = \langle \sum_{i=1}^{n} a_i K_{x_i}, \sum_{j=1}^{n} a_j K_{x_j} \rangle > 0.$$

The symmetry of the matrix $\boldsymbol{K}_X$ follows from the symmetry of K. □

Using this, we will now prove the following interpolation theorem.

Theorem 3.2.2. *Problem 1 has a unique solution that has the form*

$$s(\,\cdot\,) = \sum_{i=1}^{n} \lambda_i K_{x_i}(\,\cdot\,),$$

where $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_n) \in \mathbb{R}^n$ *is the solution to the following linear problem*

$$\boldsymbol{K}_X \boldsymbol{\lambda} = \boldsymbol{d}.$$

Proof. As $\boldsymbol{K}_X$ is positive definite the linear equation $\boldsymbol{K}_X \boldsymbol{\lambda} = \boldsymbol{d}$ has a unique solution. Let's show that

$$s(\,\cdot\,) = \sum_{i=1}^{n} \lambda_i K_{x_i}(\,\cdot\,)$$

is a solution to the minimisation problem. First let's show that it satisfies the constraints.

$$s(x_k) = \sum_{i=1}^{n} \lambda_i K_{x_i}(x_k) = (\boldsymbol{K}_X)_k \boldsymbol{\lambda} = d_k.$$

where $(\boldsymbol{K}_X)_k$ is the k^{th} row vector of $\boldsymbol{K}_X$. To show that s minimises the norm with respect to the interpolation constraints, let $s' \in H$ be such that $s'(x_i) = d_i, \ \ \forall i = 1 \ldots n$. So s' satisfies the interpolation constraints. Need to show that $\|s\| \leq \|s'\|$. As

$$\langle s' - s, s \rangle = \langle s' - s, \sum_{i=1}^{n} \lambda_i K_{x_i} \rangle = \sum_{i=1}^{n} \lambda_i (s' - s)(x_i) = \sum_{i=1}^{n} \lambda_i (d_i - d_i) = 0,$$

the following holds

$$\|s'\|^2 = \|s' - s + s\|^2 = \|s' - s\|^2 + \|s\|^2 + 2\langle s' - s, s \rangle = \|s\|^2 + \|s - s'\|^2.$$

$\|s - s'\|^2 \geq 0$, so $\|s'\|^2 \geq \|s\|^2$. This shows that s minimises the norm subject to the given constraints. For uniqueness, Suppose $\|s'\| = \|s\|$. Rearranging the equation above gives $\|s - s'\|^2 = \|s'\|^2 - \|s\|^2$. As $\|s'\|^2 - \|s\|^2 = 0$, $s' = s$. □

The above proof shows that the above minimisation problem is reducible to a linear algebraic problem. The above theorem can be generalised to interpolation with respect to continuous linear functionals over H.

3.2.2 Interpolation with continuous linear functionals

Consider a variant of Problem 1.

Problem 2. *Let $(\Omega, H, \langle\cdot,\cdot\rangle, K)$ be a RKHS. Let $\mathcal{X} = \{\mu_1, \mu_2, \ldots, \mu_n\} \subseteq H^*$ be a set of $n \in \mathbb{N}$ linearly independent functionals in H^*. Find $s \in H$ such that*

$$s = \underset{f \in H}{\arg\min} \|f\|^2 \text{ subject to } \mu_i(f) = d_i, \ \ \forall i = 1 \ldots n \tag{5}$$

where $\boldsymbol{d} = (d_1, d_2, \ldots, d_n) \in \mathbb{R}^n$.

Note the differences between Problem 1 and Problem 2. in Problem 1, K is strictly positive definite. This criteria has been replaced with requiring that $\mathcal{X}$ is a set of linearly independent functionals in H^*. If we required K to be strictly positive definite and made $\mathcal{X}$ to be a set of point evaluation functionals, then it would follow from Theorem 3.1.5 that $\mathcal{X}$ is linearly independent.

Problem 2 has a unique solution. Before proving this let's prove some basic facts concerning the relationship between reproducing kernels and their corresponding continous linear functionals. Let $(\Omega, H, \langle\cdot,\cdot\rangle, K)$ be a RKHS and let $\mu \in H^*$. By the Riesz representation theorem, we have $\mu(f) = \langle f, R(\mu)\rangle$ where $R : H^* \to H$ is the representation map between H^* and H. $R(\mu)$ is related to K as follows.

Lemma 3.2.3. *Let $(\Omega, H, \langle\cdot,\cdot\rangle, K)$ be a RKHS. Let $\mu \in H^*$. $R(\mu) = \mu^{(1)}K = \mu^{(2)}K$.*

Proof. Let $x \in \Omega$.

$$R(\mu)(x) = \langle R(\mu), K_x\rangle = \langle K_x, R(\mu)\rangle = \mu^{(2)}K(x,\cdot).$$

Hence $R(\mu) = \mu^{(2)}K$. By the symmetry of K,

$$\mu^{(2)}K(x,\cdot) = \mu^{(1)}K(\cdot,x)$$

for any $x \in \Omega$. So $\mu^{(1)}K = \mu^{(2)}K$. □

Given this, there is no ambiguity in writing μK, $\forall \mu \in H^*$.

Corollary 3.2.3.1. *Let $(\Omega, H, \langle\cdot,\cdot\rangle, K)$ be a RKHS. For any $\mu_1, \mu_2 \in H^*$. $\mu_1^{(1)}\mu_2^{(2)}K = \mu_2^{(1)}\mu_1^{(2)}K$.*

Proof. As $\mu_1 K = R(\mu_1)$ and $\mu_2 K = R(\mu_2)$, we have the following:

$$\begin{aligned}\mu_1^{(1)}\mu_2^{(2)}K = \mu_1 R(\mu_2) = \langle R(\mu_2), R(\mu_1)\rangle = \langle R(\mu_1), R(\mu_2)\rangle &= \mu_2 R(\mu_1) \\ &= \mu_2^{(1)}\mu_1^{(2)}K.\end{aligned}$$

Hence $\mu_1^{(1)}\mu_2^{(2)}K = \mu_2^{(1)}\mu_1^{(2)}K$. □

Given this corollary, there is no ambiguity in writing $\mu_1\mu_2 K$ or μK. Letting $\mathcal{X} = \{\mu_1, \mu_2, \ldots, \mu_n\} \subset H^*$ be $n \in \mathbb{N}$ distinct continuous linear functionals in H^*, the $n \times n$ matrix $\boldsymbol{K}_{\mathcal{X}}$ defined as follows

$$(\boldsymbol{K}_{\mathcal{X}})_{ij} = \mu_i\mu_j K,$$

is the **Gramian matrix** of K at $\mathcal{X}$.

Lemma 3.2.4. *Let $\mathcal{X} = \{\mu_1, \mu_2, \ldots, \mu_n\} \subseteq H^*$ be $n \in \mathbb{N}$ distinct linearly independent continuous linear functionals in H^*. $\boldsymbol{K}_{\mathcal{X}}$ is symmetric positive definite.*

Proof. Symmetric follows from Corollary 3.2.3.1. Let $\boldsymbol{a} = (a_1, a_2, \ldots, a_n) \in \mathbb{R}^n \backslash \{\mathbf{0}\}$.

$$\begin{aligned}\boldsymbol{a}^{\mathbf{T}}\boldsymbol{K}_X\boldsymbol{a} = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \mu_i\mu_j K &= \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \langle R(\mu_i), R(\mu_j)\rangle \\ &= \langle \sum_{i=1}^{n} a_i R(\mu_i), \sum_{j=1}^{n} a_j R(\mu_j)\rangle\end{aligned}$$

which is greater than zero as $\mathcal{X}$ is linearly independent in H^*. □

We will now prove the following general interpolation problem.

Theorem 3.2.5. *Problem 2 has a unique solution that has the form*

$$s(\,\cdot\,) = \sum_{i=1}^{n} \lambda_i \mu_i K(\,\cdot\,)$$

where $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_n) \in \mathbb{R}^n$ is the solution to the following linear problem

$$\boldsymbol{K}_{\mathcal{X}}\boldsymbol{\lambda} = \boldsymbol{d}.$$

Proof. This proof will mirror the previous interpolation theorem. As $\boldsymbol{K}_{\mathcal{X}}$ is positive definite the linear equation $\boldsymbol{K}_{\mathcal{X}}\boldsymbol{\lambda} = \boldsymbol{d}$ has a unique solution. Let's show that

$$s(\,\cdot\,) = \sum_{i=1}^{n} \lambda_i \mu_i K(\,\cdot\,)$$

is a solution to the minimisation problem. First let's show that it satisfies the constraints.

$$\mu_k(s) = \sum_{i=1}^{n} \lambda_i \mu_k \mu_i K = (\boldsymbol{K}_{\mathcal{X}})_k \boldsymbol{\lambda} = d_k.$$

where $(\boldsymbol{K}_{\mathcal{X}})_k$ is the k^{th} row vector of $\boldsymbol{K}_{\mathcal{X}}$. To show that s minimises the norm, let $s' \in H$ such that $\mu_i(s') = f_i, \;\; \forall i = 1 \ldots n.$ as

$$\langle s' - s, s \rangle = \sum_{i=1}^{n} a_i \mu_i(s' - s) = \sum_{i=1}^{n} a_i(d_i - d_i) = 0,$$

the following holds

$$\|s'\|^2 = \|s' - s + s\|^2 = \|s' - s\|^2 + \|s\|^2 + 2\langle s' - s, s\rangle = \|s\|^2 + \|s - s'\|^2.$$

$\|s - s'\|^2 \geq 0$, so $\|s'\|^2 \geq \|s\|^2$. This shows that s minimises the norm subject to the given constraints. For uniqueness, Suppose $\|s'\| = \|s\|$. Rearranging the equation above gives $\|s - s'\|^2 = \|s'\|^2 - \|s\|^2$. As $\|s'\|^2 - \|s\|^2 = 0$, $s' = s$. $\square$

Before the end of this subsection, let's note the following, the above theorem motivates defining the following. Let $\mathcal{X} = \{\mu_1, \mu_2, \ldots, \mu_n\} \subseteq H^*$ be a set of linearly independent functionals in H^*. From the above theorem, every interpolant over the functionals in $\mathcal{X}$ is in the span of the set $\{\mu_i K \mid i = 1, \ldots, n.\}$. Call this subspace the **approximation subspace** of $\mathcal{X}$ and denote it by $\mathcal{K}_{\mathcal{X}}$.

3.2.3 Penalised Least Squares

Now let's consider a smoothing problem. Let $(\Omega, H, \langle\cdot,\cdot\rangle, K)$ be a RKHS. Let $\mathcal{X} = \{\mu_1, \mu_2, \ldots, \mu_n\} \subseteq H^*$ a set of n linearly independent continuous functionals in H^* and $\boldsymbol{w} = (w_1, w_2, \ldots, w_n) \in \mathbb{R}^n, \boldsymbol{w} > \boldsymbol{0}$ be positive weights. Let the semi inner product $[\cdot,\cdot]_{\mathcal{X}}^{\boldsymbol{w}}$ over H be defined as follows:

$$[f, g]_{\mathcal{X}}^{\boldsymbol{w}} = \sum_{i=1}^{n} w_i \mu_i(f) \mu_i(g) \;\; \forall f, g \in H.$$

Let $\boldsymbol{d} = (d_1, d_2, \ldots, d_n) \in \mathbb{R}^n$, by Theorem 3.2.5, there exists an $s \in H$ such that $\mu_i s = d_i, \; \forall i = 1, \ldots, n$. Let $E_{\mathcal{X}}^{\boldsymbol{w},\boldsymbol{d}} : H \to \mathbb{R}$ be defined as follows:

$$\begin{aligned} E_{\mathcal{X}}^{\boldsymbol{w},\boldsymbol{d}}(f) &= \langle f, f\rangle + [f - s, f - s]_{\mathcal{X}}^{\boldsymbol{w}} \\ &= \langle f, f\rangle + \sum_{i=1}^{n} w_i(\mu_i(f) - d_i)^2, \;\; \forall f \in H. \end{aligned}$$

Call this function the **energy** function. The penalised least squares problem is as follows.

Problem 3. *Let $(\Omega, H, \langle\cdot,\cdot\rangle, K)$ be a RKHS. Let $\mathcal{X} = \{\mu_1, \mu_2, \ldots, \mu_n\} \subseteq H^*$ be a set of $n \in \mathbb{N}$ linearly independent functionals in H^*, $\boldsymbol{w} = (w_1, w_2, \ldots, w_n) \in \mathbb{R}^n, \boldsymbol{w} > \boldsymbol{0}$, $\boldsymbol{d} = (d_1, d_2, \ldots, d_n) \in \mathbb{R}^n$. Define the following function*

$$E_{\mathcal{X}}^{\boldsymbol{w},\boldsymbol{d}}(f) = \langle f, f\rangle + \sum_{i=1}^{n} w_i(\mu_i(f) - d_i)^2, \;\; \forall f \in H.$$

Find $s \in H$ such that

$$s = \operatorname*{arg\,min}_{f \in H} E_{\mathcal{X}}^{\boldsymbol{w},\boldsymbol{d}}(f). \tag{6}$$

This problem has a unique solution. To prove this, it will be utilised that $E_{\mathcal{X}}^{\boldsymbol{w},\boldsymbol{d}}$ is a continuous, strictly convex function. See Appendix B for the nessesary results concerning convex functions and sets.

Theorem 3.2.6. *Let $(\Omega, H, \langle\cdot,\cdot\rangle, K)$ be a RKHS. The function $E := E_{\mathcal{X}}^{\boldsymbol{w},\boldsymbol{d}}$, defined as above, is continuous.*

Proof. Trivially $\langle f, f\rangle$, is continous. As μ_i is in H^*, the function $\sum_{i=1}^{n} w_i(\mu_i(f) - d_i)^2$ is continuous. Thus E is continuous. $\square$

Theorem 3.2.7. *Let $(\Omega, H, \langle\cdot,\cdot\rangle, K)$ be a RKHS. The function $E := E_{X}^{\boldsymbol{w},\boldsymbol{d}}$, defined as above, is strictly convex.*

Proof. This follows from the fact that the functions $f \to \langle f, f\rangle$ and $f \to [f - s, f - s]_{\mathcal{X}}^{\boldsymbol{w}}$ are both strictly convex by theorem B.0.3 and theorem B.0.1, hence their sum is strictly convex by theorem B.0.2. $\square$

Theorem 3.2.8. *Let $(\Omega, H, \langle\cdot,\cdot\rangle, K)$ be a RKHS. and let $E := E_{\mathcal{X}}^{\boldsymbol{w},\boldsymbol{d}}$ be defined as above. There exists a unique minimiser of E over H.*

Proof. Let's first outline this proof. A compact convex subset of H will be found such that if there is a local minimiser of E in this compact convex set, then said local minimiser would be a global minimiser of E over H. Utilising the fact that E is strictly convex, said local minimum must be unique by Theorem B.0.6. Hence there is a unique global minimiser of E over H.

Let $f \in H$. From Theorem 3.2.5, there exists unique minimal norm interpolant of f (say If) such that $\mu_i(If) = \mu_i(f)$ for all $i = 1, \dots, n$. If is in the approximation subspace $\mathcal{K}_{\mathcal{X}}$ of $\mathcal{X}$. $E(f) - E(If) = \langle f, f\rangle - \langle If, If\rangle \geq 0$ with equality only when $f = If$ so if a minimiser of E exists, it must exist in $\mathcal{K}_{\mathcal{X}}$. Let $s \in \mathcal{K}_{\mathcal{X}}$ such that $\mu_i s = d_i$, $\forall i = 1, \dots, n$. It follows that $E(s) = \langle s, s\rangle$. Now consider the set $\{f \in \mathcal{K}_{\mathcal{X}} \mid E(f) \leq E(s) = \langle s, s\rangle\}$. This set is a closed subset of the set $\{f \in \mathcal{K}_{\mathcal{X}} \mid \langle f, f\rangle \leq \langle s, s\rangle\}$ which is compact as it is bounded, closed and finite dimensional (as $\mathcal{K}_{\mathcal{X}}$ is of dimension n). It follows that $\{f \in \mathcal{K}_{\mathcal{X}} \mid E(f) \leq E(s)\}$ is also compact. As E is continuous over H, it achieves a minimum on $\{f \in \mathcal{K}_{\mathcal{X}} \mid E(f) \leq E(s)\}$. This minimum is a local minimum because if every neighborhood around the minimum contained a function of a lower energy evaluation, then said function would be in $\{f \in \mathcal{K}_{\mathcal{X}} \mid E(f) \leq E(s)\}$, which is a contradiction.

For the uniqueness of this minimiser, lets show that this set is not just compact, but convex. Recall first the strict convexity of E established in Theorem 3.2.7. Let $f_1, f_2 \in \{f \in \mathcal{K}_{\mathcal{X}} \mid E(f) \leq E(s)\}$. Let $\theta \in (0, 1)$,

$$E(\theta f_1 + (1 - \theta)f_2) \leq \theta E(f_1) + (1 - \theta)E(f_2) \leq \theta E(s) + (1 - \theta)E(s) = E(s).$$

Thus $\theta f_1 + (1-\theta)f_2 \in \{f \in \mathcal{K}_{\mathcal{X}} \mid E(f) \leq E(s)\}$. As E is strictly convex over H and achieves a minimum in H, said minimum is unique by Theorem B.0.6. $\square$

Now let's consider finding the unique solution to minimising $E_{\mathcal{X}}^{\boldsymbol{w},\boldsymbol{d}}$. The unique solution (say $s_{\mathcal{X}}^{\boldsymbol{w},\boldsymbol{d}}$) to this problem is of the form

$$s_{\mathcal{X}}^{\boldsymbol{w},\boldsymbol{d}} = \sum_{i=1}^{n} a_i \mu_i K$$

where $\boldsymbol{a} = (a_1, a_2, \ldots, a_n) \in \mathbb{R}^n$. Let $\boldsymbol{W}$ is the $n \times n$ diagonal matrix with diagonal entries $(\boldsymbol{W})_{ii} = w_i \;\; \forall i = 1, \ldots, n$. Inputting $s_{\mathcal{X}}^{W}$ into $E_{\mathcal{X}}^{W}$ gives:

$$\begin{aligned}
E_{\mathcal{X}}^{\boldsymbol{w},\boldsymbol{d}}(s_{\mathcal{X}}^{\boldsymbol{w},\boldsymbol{d}}) &= \langle s_{\mathcal{X}}^{\boldsymbol{w},\boldsymbol{d}}, s_{\mathcal{X}}^{\boldsymbol{w},\boldsymbol{d}} \rangle + \sum_{i=1}^{n} w_i (\mu_i(s_{\mathcal{X}}^{\boldsymbol{w},\boldsymbol{d}}) - d_i)^2 \\
&= \sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j \mu_j^{(2)} \mu_i^{(1)} K + \sum_{k=1}^{n} w_k (\sum_{l=1}^{n} a_l \mu_k^{(2)} \mu_l^{(1)} K - d_k)^2 \\
&= \boldsymbol{a}^T \boldsymbol{K}_{\mathcal{X}} \boldsymbol{a} + (\boldsymbol{K}_{\mathcal{X}} \boldsymbol{a} - \boldsymbol{d})^T \boldsymbol{W} (\boldsymbol{K}_{\mathcal{X}} \boldsymbol{a} - \boldsymbol{d}) \\
&= \boldsymbol{a}^T \boldsymbol{K}_{\mathcal{X}} \boldsymbol{a} + \|\boldsymbol{K}_{\mathcal{X}} \boldsymbol{a} - \boldsymbol{d}\|_{\boldsymbol{w}}^2,
\end{aligned}$$

where $\|\cdot\|_{\boldsymbol{w}} : \mathbb{R}^n \to \mathbb{R}$ is the norm $\|\boldsymbol{v}\|_{\boldsymbol{w}} = \sqrt{\boldsymbol{v}^t \boldsymbol{W} \boldsymbol{v}}$.
Define $\mathcal{E}_{\mathcal{X}}^{\boldsymbol{w},\boldsymbol{d}} : \mathbb{R}^n \to \mathbb{R}$ as follows:

$$\mathcal{E}_{\mathcal{X}}^{\boldsymbol{w},\boldsymbol{d}}(\boldsymbol{c}) = \boldsymbol{c}^T \boldsymbol{K}_{\mathcal{X}} \boldsymbol{c} + \|\boldsymbol{K}_{\mathcal{X}} \boldsymbol{c} - \boldsymbol{d}\|_{\boldsymbol{w}}^2, \;\; \forall \boldsymbol{c} \in \mathbb{R}^n.$$

Given the equivalence of $\mathcal{E}_{\mathcal{X}}^{\boldsymbol{w},\boldsymbol{d}}$ over $\mathbb{R}^n$ and $E_{\mathcal{X}}^{\boldsymbol{w},\boldsymbol{d}}$ over V, there exists a unique local minimiser of $\mathcal{E}_{\mathcal{X}}^{W}$ over $\mathbb{R}^n$ (which is $\boldsymbol{a}$ from before). As the gradient of $\mathcal{E}_{\mathcal{X}}^{W}$ exists, the solution to minimisation problem satisfies the following equation:

$$\nabla \mathcal{E}_{\mathcal{X}}^{\boldsymbol{w},\boldsymbol{d}}(\boldsymbol{a}) = \boldsymbol{0}.$$

From this

$$\begin{aligned}
\nabla \mathcal{E}_{\mathcal{X}}^{\boldsymbol{w},\boldsymbol{d}}(\boldsymbol{a}) = 2\boldsymbol{K}_{\mathcal{X}} \boldsymbol{a} + 2\boldsymbol{K}_{\mathcal{X}} \boldsymbol{W} \boldsymbol{K}_{\mathcal{X}} \boldsymbol{a} - 2\boldsymbol{K}_{\mathcal{X}} \boldsymbol{W} \boldsymbol{d} &= 0. \\
\implies 2\boldsymbol{K}_{\mathcal{X}} \boldsymbol{a} + 2\boldsymbol{K}_{\mathcal{X}} \boldsymbol{W} \boldsymbol{K}_{\mathcal{X}} \boldsymbol{a} &= 2\boldsymbol{K}_{\mathcal{X}} \boldsymbol{W} \boldsymbol{d}. \\
\implies \boldsymbol{a} + \boldsymbol{W} \boldsymbol{K}_{\mathcal{X}} \boldsymbol{a} &= \boldsymbol{W} \boldsymbol{d}. \\
\implies (\boldsymbol{W}^{-1} + \boldsymbol{K}_{\mathcal{X}}) \boldsymbol{a} &= \boldsymbol{d}.
\end{aligned}$$

The following theorem has been established:

Theorem 3.2.9. *Problem 3 has a unique solution that has the form*

$$s(\,\cdot\,) = \sum_{i=1}^{n} \lambda_i \mu_i K(\,\cdot\,),$$

where $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_n) \in \mathbb{R}^n$ *is the solution to the following linear problem*

$$(\boldsymbol{W}^{-1} + \boldsymbol{K}_{\mathcal{X}}) \boldsymbol{\lambda} = \boldsymbol{d},$$

and $\boldsymbol{W}$ *is the* $n \times n$ *diagonal matrix with diagonal entries* $(\boldsymbol{W})_{ii} = w_i \;\; \forall i = 1, \ldots, n$.

3.2.4 Mixed Interpolation and Smoothing

Finally, let's consider a mixed interpolation and smoothing problem. Let $\mathcal{X} = \{\mu_1, \mu_2, \dots, \mu_n\} \subseteq H^*$ a set of n linearly independent continuous functionals in H^* and $\boldsymbol{w} = (w_1, w_2, \dots, w_N) \in \mathbb{R}^N, \boldsymbol{w} > \mathbf{0}$, $N < n$. Now let $\mathcal{Y} = \{\mu_1, \mu_2, \dots, \mu_N\} \subset \mathcal{X}$. Let the semi inner product $[\cdot,\cdot]_{\mathcal{Y}}^{\boldsymbol{w}}$ over H be defined as follows:

$$[f, g]_{\mathcal{Y}}^{\boldsymbol{w}} = \sum_{i=1}^{N} w_i \mu_i(f)\mu_i(g) \;\; \forall f, g \in H.$$

Let $\boldsymbol{d} = (d_1, d_2, \dots, d_n) \in \mathbb{R}^n$, by Theorem 3.2.5, there exists an $s \in H$ such that $\mu_i s = d_i$, $\forall i = 1, \dots, n$. Let $E_{\mathcal{Y}}^{\boldsymbol{w},\boldsymbol{d}} : H \to \mathbb{R}$ be defined as follows:

$$\begin{aligned} E_{\mathcal{Y}}^{\boldsymbol{w},\boldsymbol{d}}(f) &= \langle f, f\rangle + [f - s, f - s]_{\mathcal{Y}}^{\boldsymbol{w}} \\ &= \langle f, f\rangle + \sum_{i=1}^{N} w_i(\mu_i(f) - d_i)^2, \;\; \forall f \in H. \end{aligned}$$

The mixed interpolation and smoothing problem is as follows.

Problem 4. *Let $(\Omega, H, \langle\cdot,\cdot\rangle, K)$ be a RKHS. Let $\mathcal{X} = \{\mu_1, \mu_2, \dots, \mu_n\} \subseteq H^*$ be a set of $n \in \mathbb{N}$ linearly independent functionals in H^*, $\boldsymbol{w} = (w_1, w_2, \dots, w_n) \in \mathbb{R}^N, \boldsymbol{w} > \mathbf{0}$, $N < n$, $\boldsymbol{d} = (d_1, d_2, \dots, d_n) \in \mathbb{R}^n$. Define the following function*

$$E_{\mathcal{X}}^{\boldsymbol{w},\boldsymbol{d}}(f) = \langle f, f\rangle + \sum_{i=1}^{N} w_i(\mu_i(f) - d_i)^2, \;\; \forall f \in H.$$

Find $s \in H$ such that

$$s = \arg\min_{f \in H} E_{\mathcal{Y}}^{\boldsymbol{w},\boldsymbol{d}}(f). \tag{7}$$

It will be shown that there is a unique solution to this problem in H.

Theorem 3.2.10. *Let $(\Omega, H, \langle\cdot,\cdot\rangle, K)$ be a RKHS. The function $E := E_{\mathcal{Y}}^{\boldsymbol{w},\boldsymbol{d}}$, defined as above, is continuous.*

Proof. Trivially $\langle f, f\rangle$, is continous. As μ_i is in H^*, the function $\sum_{i=1}^{N} w_i(\mu_i(f) - d_i)^2$ is continuous. Thus E is continuous. □

Theorem 3.2.11. *Let $(\Omega, H, \langle\cdot,\cdot\rangle, K)$ be a RKHS. The function $E := E_{\mathcal{Y}}^{\boldsymbol{w},\boldsymbol{d}}$, defined as above, is strictly convex.*

Proof. This follows from the fact that the functions $f \to \langle f, f\rangle$ and $f \to [f - s, f - s]_{\mathcal{Y}}^{\boldsymbol{w}}$ are both strictly convex, hence their sum is strictly convex. □

Theorem 3.2.12. *Let $(\Omega, H, \langle\cdot,\cdot\rangle, K)$ be a RKHS and let $E_{\mathcal{Y}}^{\boldsymbol{w},\boldsymbol{d}} =: E$ be defined as above. there exists a unique minimiser of E over H.*

Proof. Let $C_{\mathcal{X}}$ be the set of all functions that satisfy the interpolation constraints, this set is convex and closed. Let $f \in C_{\mathcal{X}}$. From Theorem 3.2.5, there exists unique minimal norm interpolant of f (say If) such that $\mu_i(If) = \mu_i(f)$ for all $i = 1, \dots, n$. If is in the approximation subspace $\mathcal{K}_{\mathcal{X}} = \text{span}\{\mu_i K \mid i = 1, \dots, n.\}$. $E(f) \geq E(If)$ with equality only when $f = If$ and so if there exists a minimiser of E over H subject to the interpolation constraints, then they exist in $\mathcal{K}_{\mathcal{X}} \cap C_{\mathcal{X}}$. Let $s \in \mathcal{K}_{\mathcal{X}}$ such that $\mu_i(s) = d_i, \ \forall i = 1, \dots, n$, it follows that $E(s) = \langle s, s \rangle$. Now consider the set $\{f \in \mathcal{K}_{\mathcal{X}} \mid E(f) \leq E(s) = \langle s, s \rangle\}$. This set is a closed subset of the set $\{f \in \mathcal{K}_{\mathcal{X}} \mid \langle f, f \rangle \leq \langle s, s \rangle\}$ which is compact as it is bounded, closed and finite dimensional (As $\mathcal{K}_{\mathcal{X}}$ is of dimension n). It follows that $\{f \in \mathcal{K}_{\mathcal{X}} \mid E(f) \leq E(s)\} \cap C_{\mathcal{X}}$ is also compact. As E is continuous over H, it achieves a minimum on $\{f \in \mathcal{K}_{\mathcal{X}} \mid E(f) \leq E(s)\} \cap C_{\mathcal{X}}$. This minimum is a local minimum (within $C_{\mathcal{X}}$) because if every neighborhood around the minimum contained a function that satisfied the interpolation constraints yet had a lower energy evaluation, then said function would be in $\{f \in \mathcal{K}_{\mathcal{X}} \mid E(f) \leq E(s)\} \cap C_{\mathcal{X}}$, which is a contradiction.

Toward the uniqueness of this minimiser, let us establish that the set $\{f \in \mathcal{K}_{\mathcal{X}} \mid E(f) \leq E(s)\}$ is convex. Recall form Theorem 3.2.10 and Theorem 3.2.11 that E is a strictly convex function. Let $f_1, f_2 \in \{f \in \mathcal{K}_{\mathcal{X}} \mid E(f) \leq E(s)\}$. Let $\theta \in (0, 1)$,

$$E(\theta f_1 + (1 - \theta) f_2) \leq \theta E(f_1) + (1 - \theta) E(f_2) \leq \theta E(s) + (1 - \theta) E(s) = E(s).$$

Thus $\theta f_1 + (1 - \theta) f_2 \in \{f \in \mathcal{K}_{\mathcal{X}} \mid E(f) \leq E(s)\}$. It follows that $\{f \in \mathcal{K}_{\mathcal{X}} \mid E(f) \leq E(s)\} \cap C_{\mathcal{X}}$ is convex as $C_{\mathcal{X}}$ is convex. As E is strictly convex over $\{f \in \mathcal{K}_{\mathcal{X}} \mid E(f) \leq E(s)\} \cap C_{\mathcal{X}}$ and achieves a local minimum in $C_{\mathcal{X}}$ by Theorem B.0.6, said minimum is unique. As all minimisers of E subject to the interpolation constraints must exist in $\{f \in \mathcal{K}_{\mathcal{X}} \mid E(f) \leq E(s)\} \cap C_{\mathcal{X}}$, the minimiser is unique on the whole of H. $\square$

Now let's consider finding the unique solution to minimising $E_{\mathcal{Y}}^{\boldsymbol{w},\boldsymbol{d}}$. The unique solution (say $s_{\mathcal{Y}}^{\boldsymbol{w},\boldsymbol{d}}$) to this problem is of the form

$$s_{\mathcal{Y}}^{\boldsymbol{w},\boldsymbol{d}} = \sum_{i=1}^{n} a_i \mu_i K$$

where $\boldsymbol{a} = (a_1, a_2, \dots, a_n) \in \mathbb{R}^n$. Let $\boldsymbol{W}$ is the $n \times n$ diagonal matrix with diagonal entries $(\boldsymbol{W})_{ii} = w_i \ \ \forall i = 1, \dots, N$, $(\boldsymbol{W})_{ii} = 0, \ \forall i = N + 1, \dots, n$.

Substituting $s_{\mathcal{Y}}^{W}$ into $E_{\mathcal{Y}}^{W}$ gives:

$$\begin{aligned}
E_{\mathcal{Y}}^{\boldsymbol{w},\boldsymbol{d}}(s_{\mathcal{Y}}^{\boldsymbol{w},\boldsymbol{d}}) &= \langle s_{\mathcal{Y}}^{\boldsymbol{w},\boldsymbol{d}}, s_{\mathcal{Y}}^{\boldsymbol{w},\boldsymbol{d}}\rangle + \sum_{i=1}^{N} w_i(\mu_i(s_{\mathcal{Y}}^{\boldsymbol{w},\boldsymbol{d}}) - d_i)^2 \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \mu_j^{(2)}\mu_i^{(1)} K + \sum_{k=1}^{N} w_k(\sum_{l=1}^{n} a_l \mu_k^{(2)}\mu_l^{(1)} K - d_k)^2 \\
&= \boldsymbol{a}^T \boldsymbol{K}_{\mathcal{X}}\boldsymbol{a} + (\boldsymbol{K}_{\mathcal{X}}\boldsymbol{a} - \boldsymbol{d})^T \boldsymbol{W}(\boldsymbol{K}_{\mathcal{X}}\boldsymbol{a} - \boldsymbol{d}) \\
&= \boldsymbol{a}^T \boldsymbol{K}_{\mathcal{X}}\boldsymbol{a} + \|\boldsymbol{K}_{\mathcal{X}}\boldsymbol{a} - \boldsymbol{d}\|_{\boldsymbol{w}}^2,
\end{aligned}$$

where $\|\cdot\|_{\boldsymbol{w}} : \mathbb{R}^n \to \mathbb{R}$ is the semi norm $\|\boldsymbol{v}\|_{\boldsymbol{w}} = \sqrt{\boldsymbol{v}^t \boldsymbol{W} \boldsymbol{v}}$.
Define $\mathcal{E}_{\mathcal{X}}^{\boldsymbol{w},\boldsymbol{d}} : \mathbb{R}^n \to \mathbb{R}$ as follows:

$$\mathcal{E}_{\mathcal{X}}^{\boldsymbol{w},\boldsymbol{d}}(\boldsymbol{c}) = \boldsymbol{c}^T \boldsymbol{K}_{\mathcal{X}}\boldsymbol{c} + \|\boldsymbol{K}_{\mathcal{X}}\boldsymbol{c} - \boldsymbol{d}\|_{\boldsymbol{w}}^2, \ \ \forall \boldsymbol{c} \in \mathbb{R}^n.$$

Given the equivalence of $\mathcal{E}_{\mathcal{X}}^{\boldsymbol{w},\boldsymbol{d}}$ over $\mathbb{R}^n$ and $E_{\mathcal{X}}^{\boldsymbol{w},\boldsymbol{d}}$ over V, there exists a unique minimiser of $\mathcal{E}_{\mathcal{X}}^{W}$ over $\mathbb{R}^n$ satisfying the interpolation constraints (which is $\boldsymbol{a}$ from before). The interpolation constraints can be expressed in the following convenient form

$$\boldsymbol{D}(\boldsymbol{K}_{\mathcal{X}}\boldsymbol{a} - \boldsymbol{d}) = \boldsymbol{0},$$

or equivalently

$$\boldsymbol{D}\boldsymbol{K}_{\mathcal{X}}\boldsymbol{a} = \boldsymbol{D}\boldsymbol{d},$$

where $\boldsymbol{D}$ is the $n \times n$ diagonal matrix such that $\boldsymbol{D}_{ii} = 0, \ i = 1, \ldots, N \ \boldsymbol{D}_{ii} = 1, \ i = N+1, \ldots, n$. It is important to note the relationship between $\boldsymbol{W}$ and $\boldsymbol{D}$ given by

$$\boldsymbol{W}\boldsymbol{W}^{\dagger} = \boldsymbol{I}_n - \boldsymbol{D}, \ \boldsymbol{W}\boldsymbol{D} = \boldsymbol{W}^{\dagger}\boldsymbol{D} = \boldsymbol{0},$$

where $\boldsymbol{W}^{\dagger}$ is the pseudo inverse of $\boldsymbol{W}$ and $\boldsymbol{I}_n$ is the $n \times n$ identity matrix. As the gradient of $\mathcal{E}_{\mathcal{X}}^{W}$ exists, via the method of Lagrange multipliers:

$$\mathcal{E}_{\mathcal{X}}^{\boldsymbol{w},\boldsymbol{d}}(\boldsymbol{a}) - \nabla_{\boldsymbol{a}}\boldsymbol{\lambda}^T \boldsymbol{D}(\boldsymbol{K}_{\mathcal{X}}\boldsymbol{a} - \boldsymbol{d}) = \boldsymbol{0},$$

where $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_n) \in \mathbb{R}^{n-N}, \ \lambda_i = 0, \ \forall i = 1, \ldots, N$, so $\boldsymbol{D}\boldsymbol{\lambda} = \boldsymbol{\lambda}$. From this

$$\begin{aligned}
2\boldsymbol{K}_{\mathcal{X}}\boldsymbol{a} + 2\boldsymbol{K}_{\mathcal{X}}\boldsymbol{W}\boldsymbol{K}_{\mathcal{X}}\boldsymbol{a} - 2\boldsymbol{K}_{\mathcal{X}}\boldsymbol{W}\boldsymbol{d} - \boldsymbol{K}_{\mathcal{X}}\boldsymbol{\lambda} &= \boldsymbol{0}, \\
\implies 2\boldsymbol{a} + 2\boldsymbol{W}\boldsymbol{K}_{\mathcal{X}}\boldsymbol{a} - 2\boldsymbol{W}\boldsymbol{d} &= \boldsymbol{\lambda}.
\end{aligned}$$

Left multiplying by $\boldsymbol{W}^{\dagger}$, the pseudo inverse of $\boldsymbol{W}$, gives

$$\begin{aligned}
2\boldsymbol{W}^{\dagger}\boldsymbol{a} + 2\boldsymbol{W}^{\dagger}\boldsymbol{W}\boldsymbol{K}_{\mathcal{X}}\boldsymbol{a} - 2\boldsymbol{W}^{\dagger}\boldsymbol{W}\boldsymbol{d} = \boldsymbol{W}^{\dagger}\boldsymbol{\lambda} = (\boldsymbol{W}^{\dagger}\boldsymbol{D})\boldsymbol{\lambda} &= \boldsymbol{0}, \\
\implies (\boldsymbol{W}^{\dagger} + \boldsymbol{W}^{\dagger}\boldsymbol{W}\boldsymbol{K}_{\mathcal{X}})\boldsymbol{a} &= \boldsymbol{W}^{\dagger}\boldsymbol{W}\boldsymbol{d}, \\
\implies (\boldsymbol{W}^{\dagger} + (\boldsymbol{I}_n - \boldsymbol{D})\boldsymbol{K}_{\mathcal{X}})\boldsymbol{a} &= (\boldsymbol{I}_n - \boldsymbol{D})\boldsymbol{d}.
\end{aligned}$$

Combining this equation with the interpolation constraints $\boldsymbol{D}\boldsymbol{K}_{\mathcal{X}}\boldsymbol{a} = \boldsymbol{D}\boldsymbol{d}$ gives the matrix equation

$$(\boldsymbol{W}^{\dagger} + \boldsymbol{K}_{\mathcal{X}})\boldsymbol{a} = \boldsymbol{d}.$$

$\boldsymbol{W}^{\dagger} + \boldsymbol{K}_{\mathcal{X}}$ is invertible as $\boldsymbol{K}_{\mathcal{X}}$ is positive definite and $\boldsymbol{W}^{\dagger}$ is non-negative definite. The following theorem has been established:

Theorem 3.2.13. *Consider the following minimisation problem.*

Problem 4 has a unique solution that has the form

$$s(\,\cdot\,) = \sum_{i=1}^{n} \lambda_i \mu_i K(\,\cdot\,),$$

where $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \dots, \lambda_n) \in \mathbb{R}^n$ *is the solution to the following linear problem*

$$(\boldsymbol{W}^{\dagger} + \boldsymbol{K}_{\mathcal{X}})\boldsymbol{\lambda} = \boldsymbol{d},$$

and $\boldsymbol{W}$ *is the* $n \times n$ *diagonal matrix with diagonal entries* $(\boldsymbol{W})_{ii} = w_i,\ \forall i = 1, \dots, N,\ (\boldsymbol{W})_{ii} = 0,\ \forall i = N+1, \dots, n.$

3.2.5 Interpolation reproducing constants

Finally, let's now consider the following problem. Let $(\Omega, H, \langle\cdot,\cdot\rangle, K)$ be a strict RKHS with induced norm $\|\cdot\|$. Letting $X = \{x_1, x_2, \dots, x_n\} \subseteq \Omega$ of size $n \in \mathbb{N}$ and $\boldsymbol{d} = (d_1, d_2, \dots, d_n)$, by Theorem 3.2.2, there exists a minimal norm interpolant given by

$$s(\,\cdot\,) = \sum_{i=1}^{n} \lambda_i K_{x_i}(\,\cdot\,)$$

where $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \dots, \lambda_n) \in \mathbb{R}^n$ is the solution to the following linear problem

$$\boldsymbol{K}_X \boldsymbol{\lambda} = \boldsymbol{d}.$$

So $s(x_i) = d_i,\ \forall i = 1, \dots, n$. Now consider the following. Let $k \in \mathbb{R}$ be some constant. We can find the minimal norm interpolant at $(X, \boldsymbol{d} + k)$. It would be desirable if this interpolant were of the form

$$s^*(\,\cdot\,) = s(\,\cdot\,) + k.$$

Indeed, this function satisfies the interpolation constraints and is simply a translation of s. This generally will not occur, however we can alter the interpolation problem to obtain a result like this by altering the norm $\|\cdot\|$. For convenience, suppose H does not contain the constant functions $\mathbb{R}_\Omega$. Simply adjoin the set of all constant functions $\mathbb{R}_\Omega$ (which is a one dimensional vector space) and extend the inner product $\langle\cdot,\cdot\rangle$ to $\mathbb{R}_\Omega$ by making their norms zero. This extention turns the inner product of H into a semi inner product over $H \oplus \mathbb{R}_\Omega$ with null space $\mathbb{R}_\Omega$ and quotient space isomorphic to H (see Appendix A for definition of semi inner product). The result is a semi Hilbert space (say $(H', [\cdot,\cdot])$) with null space $\mathbb{R}_\Omega$, induced semi norm $|\cdot|$, H, $\mathbb{R}_\Omega \in H'$ and $|f| = \|f\|$ for any $f \in H$.

Let $\mathcal{K}_X = \{\sum_{i=1}^n a_i K_{x_i} \in H \mid \boldsymbol{a} = (a_1, a_2, \ldots, a_n) \in \mathbb{R}^n\}$. This is the interpolation space of H. Let $C_X^{\boldsymbol{d}}$ be the set of any function $\mathcal{K}_X \oplus \mathbb{R}_\Omega$ that satisfies the interpolation constraints $(X, \boldsymbol{d}+k)$ for some $k \in \mathbb{R}$. So the set $C_X^{\boldsymbol{d}}$ contains functions that almost satisfy the interpolation constraints $(X, \boldsymbol{d})$. It will be shown that there exists a unique norm minimiser in the set $C_X^{\boldsymbol{d}}$ satisfying the interpolation constraints. Let $f \in H'$ such that f satisfies the interpolation constraints. $C_X^{\boldsymbol{d}}$ has the partitioning $\{f + \mathbb{R}_\Omega \mid f \in C_X^{\boldsymbol{d}}\}$ which is a convex compact subset of $H'/\mathbb{R}_\Omega$. So the continuous strictly convex function $|\cdot|$ obtains a unique minimiser over $\{f + \mathbb{R}_\Omega \mid f \in C_X^{\boldsymbol{d}}\}$ (say $s + \mathbb{R}_\Omega$ where $s = \sum_{i=1}^n a_i K_{x_i}$). Within $s + \mathbb{R}_\Omega$ there exists a unique function $s + c$, $c \in \mathbb{R}_\Omega$ that satisfies the interpolation constraints. This is the unique $|\cdot|$ minimiser that is subject to the interpolation constraints

$$(s+c)(x_i) = d_i, \ \forall i = 1, \ldots, n,$$

which can be represented in the matrix form

$$\begin{bmatrix} \boldsymbol{K}_X & \mathbb{1} \end{bmatrix} \begin{bmatrix} \boldsymbol{a} \\ c \end{bmatrix} = \boldsymbol{K}_X \boldsymbol{a} + c\mathbb{1} = \boldsymbol{d}.$$

Here we are treating c as simply a real number. we can express $|\cdot|^2$ in an elegant form over $\mathcal{K}_X \oplus \mathbb{R}_\Omega$, in fact

$$|\sum_{i=1}^n a_i K_{x_i} + c|^2 = \begin{bmatrix} \boldsymbol{a}^T & c \end{bmatrix} \begin{bmatrix} \boldsymbol{K}_X & \boldsymbol{0} \\ \boldsymbol{0}^T & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{a}^T \\ c \end{bmatrix} = \boldsymbol{a}^T \boldsymbol{K}_X \boldsymbol{a}, \text{ call } \boldsymbol{G}_X := \begin{bmatrix} \boldsymbol{K}_X & \boldsymbol{0} \\ \boldsymbol{0}^T & 0 \end{bmatrix}.$$

Via the method of Lagrange multipliers letting $\boldsymbol{b} = (\boldsymbol{a}, c)$

$$\nabla_{\boldsymbol{b}}(\boldsymbol{b}^T \boldsymbol{G}_X \boldsymbol{b}) - \nabla_{\boldsymbol{b}}(\boldsymbol{\lambda}^T(\begin{bmatrix} \boldsymbol{K}_X & \mathbb{1} \end{bmatrix} \boldsymbol{b} - \boldsymbol{d})) = \boldsymbol{0},$$

where $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_n) \in \mathbb{R}^n$ is the vector of Lagrange multipliers. This gives

$$2\boldsymbol{G}_X \boldsymbol{b} - \begin{bmatrix} \boldsymbol{K}_X \\ \mathbb{1}^T \end{bmatrix} \boldsymbol{\lambda} = 2 \begin{bmatrix} \boldsymbol{K}_X & \boldsymbol{0} \\ \boldsymbol{0}^T & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{a} \\ c \end{bmatrix} - \begin{bmatrix} \boldsymbol{K}_X \\ \mathbb{1}^T \end{bmatrix} \boldsymbol{K}_X \boldsymbol{\lambda} = \begin{bmatrix} 2\boldsymbol{K}_X \boldsymbol{a} - \boldsymbol{K}_X \boldsymbol{\lambda} \\ \mathbb{1}^T \boldsymbol{\lambda} \end{bmatrix} = \boldsymbol{0}.$$

From this it follows that

$$2\boldsymbol{K}_X \boldsymbol{a} = \boldsymbol{K}_X \boldsymbol{\lambda}, \ \mathbb{1}^T \boldsymbol{\lambda} = 0.$$

Allowing us to conclude that $2\boldsymbol{a} = \boldsymbol{\lambda}$ and $\mathbb{1}^T(2\boldsymbol{a}) = \mathbb{1}^T \boldsymbol{a} = 0$. So we have the constraints

$$\begin{bmatrix} \boldsymbol{K}_X & \mathbb{1} \end{bmatrix} \begin{bmatrix} \boldsymbol{a} \\ c \end{bmatrix} = \boldsymbol{d}, \ \mathbb{1}^T \boldsymbol{a} = 0.$$

Placing these together in block matrix form gives us

$$\begin{bmatrix} \boldsymbol{K}_X & \mathbb{1} \\ \mathbb{1}^T & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{a} \\ c \end{bmatrix} = \begin{bmatrix} \boldsymbol{d} \\ 0 \end{bmatrix}.$$

The above black matrix is singular. To see this, let $\boldsymbol{d} = \boldsymbol{0}$, then

$$\begin{bmatrix} \boldsymbol{a}^T & c \end{bmatrix} \begin{bmatrix} \boldsymbol{K}_X & \mathbb{1} \\ \mathbb{1}^T & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{a} \\ c \end{bmatrix} = \boldsymbol{a}^T \boldsymbol{K}_X \boldsymbol{a} + c\mathbb{1}^T \boldsymbol{a} = 0.$$

Given that $\mathbb{1}^T \boldsymbol{a} = 0$ the above equation reduces down to $\boldsymbol{a}^T \boldsymbol{K}_X \boldsymbol{a} = 0$. As $\boldsymbol{K}_X$ is positive definite, $\boldsymbol{a} = \boldsymbol{0}$. This implies that $c\mathbb{1} = \boldsymbol{0}$, giving $c = 0$. A more general treatment of this type of interpolation that reproduces some vector space of functions will be considered in Chapter 5 and Chapter 6.

3.3 Additional Properties of Reproducing Kernel Hilbert Spaces

Theorem 3.3.1. *Let $(\Omega, H, \langle \cdot, \cdot \rangle, K)$ be a RKHS. Convergence with respect to the norm implies pointwise convergence.*

Proof. Let $\{f_i\}_{i \in \mathbb{N}}$ be a sequence in H that converges to f and let $x \in \Omega$.

$$|f(x) - f_i(x)| = |\langle f - f_i, K_x \rangle| \leq \|K_x\| \|f - f_i\|.$$

Hence $f_i(x) \to f(x)$ as $i \to \infty$. □

Theorem 3.3.2. *Let $(\Omega, H, \langle \cdot, \cdot \rangle, K)$ be a RKHS. Let V be a closed subspace of H. Then both $(V, \langle \cdot, \cdot \rangle)$ and $(V^\perp, \langle \cdot, \cdot \rangle)$ are RKHS's over Ω. Furthermore, If G and $G^\perp$ are the reproducing kernels of V and $V^\perp$ respectively, then*

$$K = G + G^\perp.$$

Proof. As point evaluation is bounded in H, it is also bounded in both V and $V^\perp$ when $\langle \cdot, \cdot \rangle$ is restricted down to the spaces. So V and $V^\perp$ are RKHS's over Ω. Let G and $G^\perp$ be the reproducing kernels of V and $V^\perp$ respectively. Let $f \in H$, $f = f_1 + f_2$ where $f_1 \in V$, $f_2 \in V^\perp$. Let $x \in \Omega$, as G and $G^\perp$ are the reproducing kernels of V and $V^\perp$, $G_x, \in V$, $G_x^\perp \in V^\perp$. It follows that $G_x, G_x^\perp \in H$.

$$\begin{aligned} \langle f, G_x + G_x^\perp \rangle &= \langle f_1 + f_2, G_x + G_x^\perp \rangle \\ &= \langle f_1, G_x \rangle + \langle f_2, G_x^\perp \rangle \\ &= f_1(x) + f_2(x) = f(x). \end{aligned}$$

Hence $G + G^\perp$ is a reproducing kernel for H. As reproducing kernels for RKHS's are unique, $K = G + G^\perp$. □

Theorem 3.3.3. *Let $(\Omega, H, \langle \cdot, \cdot \rangle, K)$ be a RKHS. the set $\{K_x \in H \mid x \in \Omega\}$ is total.*

Proof. Let V be the closure of the span of $\{K_x \in H \mid x \in \Omega\}$ in H. Then $H = V \oplus V^{\perp}$. Let $f \in V^{\perp}$.

$$f(x) = \langle f, K_x \rangle = 0.$$

Hence f is the zero function, implying that $V^{\perp} = 0$ and so $H = V$. $\square$

4 Examples of Symmetric Positive Definite Functions

4.1 Radial Basis Functions on $\mathbb{R}^d$

Let $\Omega = \mathbb{R}^d$, $d \in \mathbb{N}$. Let $\Phi : \mathbb{R}^d \to \mathbb{R}$. Φ is **radial** if $\Phi(x) = \phi(\|x\|)$, where $\phi : [0, \infty) \to \mathbb{R}$.

A function $f : [0, \infty) \to [0, \infty)$ is **completely monotone** if

(1) $f \in [0, \infty)$,

(2) $f \in C^\infty(0, \infty)$,

(3) $(-1)^k f^{(k)}(t) \geq 0,\ \forall t > 0, k \in \mathbb{N}$.

The following theorem of Bernstein and Widder characterises completely monotone functions (Taken from [6]).

Theorem 4.1.1. *A function $f : [0, \infty) \to [0, \infty)$ is completly monotone if and only if there is a nondecreasing bounded function γ such that $f(t) = \int_0^\infty e^{-st} d\gamma(s)$.*

Schoenberg (1938) [12] realates completely monotone functions to radial functions as follows.

Theorem 4.1.2. *If f is completely monotone but not constant on $[0, \infty)$, then the function $x \to f(\|x\|^2)$ is radial, strictly positive definite on $\mathbb{R}^d$.*

An important example of a radial functions for interpolation are the Guassian functions given by

$$e^{-\alpha\|x\|^2},\ \alpha > 0.$$

Another class of important radial functions are the inverse multiquadrics given by

$$(c^2 + \|x\|_2^2)^{-\beta},\ c > 0, \beta > d/2.$$

Both of these classes are of strictly positive definite functions.

4.2 Zonal Functions on the Spheres

4.2.1 Properties of the Spheres

Let $d \in \mathbb{N}$. The $(d-1)$-dimensional sphere, $\mathbb{S}^{d-1}$, is the unit sphere of $\mathbb{R}^{m+1}$. i.e.

$$p = (x_1, x_2, \ldots, x_d) \in \mathbb{S}^{(d-1)} \text{ if and only if } x_1^2 + x_2^2 + \cdots + x_d^2 = 1$$

Clearly the sphere $\mathbb{S}^1$ is the unit circle of $\mathbb{R}^2$ and the sphere $\mathbb{S}^2$ is the unit sphere of $\mathbb{R}^3$. We also have the sphere $\mathbb{S}^\infty$ which is the unit sphere of ℓ^2. so we have

$$p = (x_1, x_2, \dots) \in \mathbb{S}^\infty \text{ if and only if } x_1^2 + x_2^2 + \cdots = 1$$

There is a natural metric on the spheres given by the shortest arc length between two points on the sphere. Let p and q be in $\mathbb{S}^{d-1}$ or $\mathbb{S}^\infty$,

$$p \cdot q = \|p\| \, \|q\| \, cos(\theta) = cos(\theta),$$

where $\cdot$ denotes the dot product and θ denotes the angle between p and q. As we are dealing with unit spheres, θ is not only the angle between p and q, it is also the shortest arc length between p and q. So we can define a metric d on $\mathbb{S}^{d-1}$ or $\mathbb{S}^\infty$ as follows:

$$d(p,q) = \theta = \arccos(p \cdot q).$$

Next we will consider the polar coordination of the sphere $\mathbb{S}^m$. Letting $(x_1, x_2, \dots, x_{m+1})$ be a point of $\mathbb{S}^m$ in rectangular coordinates, the corresponding polar coordinate vector $(\theta_1, \theta_2, \dots, \theta_d)$, where $0 \leq \theta_i \leq \pi$ when $i = 1, \dots, d-2$ and $0 \leq \theta_{d-1} \leq 2\pi$. is as follows:

$$\begin{aligned}
&x_1 = \cos(\theta_1)\\
&x_2 = \sin(\theta_1)\cos(\theta_2)\\
&x_3 = \sin(\theta_1)\sin(\theta_2)\cos(\theta_3)\\
&\vdots\\
&x_{d-2} = \sin(\theta_1)\sin(\theta_2)\cdots\sin(\theta_{d-3})\cos(\theta_{d-2})\\
&x_{d-1} = \sin(\theta_1)\sin(\theta_2)\cdots\sin(\theta_{d-2})\cos(\theta_{d-1})\\
&x_d = \sin(\theta_1)\sin(\theta_2)\cdots\sin(\theta_{d-2})\sin(\theta_{d-1})
\end{aligned}$$

We will be concerned with positive definite functions of the form

$$f \circ d(p,q) = \begin{cases} f(|\theta_p - \theta_q|) & \text{if } |\theta_p - \theta_q| \leq \pi, \\ f(2\pi - |\theta_p - \theta_q|) & \text{else,} \end{cases} \quad \text{where } \ f \in C[0,\pi].$$

Such a function f will be called **zonal** if $f \circ d$ is positive definite in the sense defined in Section 2.1. A fundamental early work on this topic is Schoenberg [13]. Now let's give an example of a positive definite function on the spheres. The following theorem is from Cheney & Light [6].

Theorem 4.2.1. *The function* $\cos(k\ \cdot\ )$ *where* $k \in \mathbb{N}$ *is positive definite on* $\mathbb{S}^{d-1}$

Proof. Let $p_1, \dots, p_n$ be n distinct points on $\mathbb{S}^m$. Let $a_1, \dots, a_n \in \mathbb{R}$.

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j cos(kd(p_i, pj)) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j cos(k \ \arccos(p_i \cdot pj))$$

As $cos(k\cdot)$ is even and $cos(k(2\pi-\theta)) = cos(k\theta)$, we have

$$\begin{aligned}\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j cos(kd(p_i, pj)) &= \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j cos(k(\theta_i - \theta_j)) \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j (cos(k\theta_i)cos(k\theta_j) + sin(k\theta_i)sin(k\theta_j)) \\ &= (\sum_{i=1}^{n} a_i cos(k\theta_i))^2 + (\sum_{j=1}^{n} a_j sin(k\theta_j))^2 > 0\end{aligned}$$

□

4.2.2 Orthogonal polynomials

For the classic treatment of orthogonal polynomials, see Szegő [17].Consider a finite interval $[a, b]$ of $\mathbb{R}$. Let $w \in C[a, b]$ be a weight function, positive except at countably many points. We have the corresponding weighted variant of the $\mathcal{L}^2[a, b]$ inner product given by

$$\langle f, g\rangle = \int_a^b f(x)g(x)w(x)dx \tag{8}$$

We will assume that the weight function w is such that the integrals

$$\int_a^b p(x)w(x)dx < \infty,$$

where $p(x)$ is a polynomial over $\mathbb{R}$. As such, the polynomials over R exist in the inner product space corresponding to $\langle\cdot,\cdot\rangle$. We know that the monomials $1, x, x^2, \ldots$ form a basis (the standard basis) for the polynomials over $\mathbb{R}$. As such, we can orthogonalise this basis via the Gram-Schmidt process. The interval $[a, b]$ for the integral defining the inner product (1) will be set to $[-1, 1]$.

An example of an important weight function w is

$$w(x) = (1-x)^\alpha(1+x)^\beta, \quad \alpha, \beta > -1.$$

The corresponding orthogonalised polynomials, orthogonalised via the Gram-Schmidt process, are called the **Jacobi polynomials**. They will be denoted by $P_n^{(\alpha,\beta)}$ where $n \in \mathbb{N}$ is the degree of the polynomial.

We are concerned with a class of Jacobi polynomials, specifically those polynomials where $\alpha = \beta$. These are referred to as the **Gengenbauer** (or **Ultraspherical**) polynomials and have the corresponding weight functions of the form

$$w(x) = (1-x)^\alpha(1+x)^\beta = (1-x^2)^\alpha = (1-x^2)^{\lambda-1/2}, \quad \alpha > -1,$$

where $\lambda = \alpha + 1/2$. These polynomials will be denoted by $C_n^{(\lambda)} = P_n^{(\alpha,\beta)}$.Actually these polynomials need to be normalised in order to be well defined. Schoenberg normalises them by setting $C_n^{(0)} = 1$. We have the following special cases

- For $\lambda = 0$, $C_n^{(0)} = T_n(x)$, the Chebyshev polynomials of the first kind.
- For $\lambda = 1/2$, $C_n^{(1/2)} = P_n(x)$, the Legendre polynomials.
- For $\lambda = 1$, $C_n^{(0)} = U_n(x)$, the Chebyshev polynomials of the second kind.

So we see that the Gengenbauer polynomials are a generalisation of these familiar classes of polynomials.

the following is a generating function for the Gegenbauer polynomials

$$\sum_{n=0}^{\infty} r^n C_n^{(\lambda)}(x) = (1 - 2rx + r^2)^{-\lambda}, \quad \lambda > 0.$$

4.2.3 Characterisation of positive definite functions on $\mathbb{S}^{d-1}$ and $\mathbb{S}^{\infty}$

The positive definite functions on $\mathbb{S}^{d-1}$ and $\mathbb{S}^{\infty}$ have been completely characterised by Schoenberg [13] in terms of Gengenbauer polynomials.

Theorem 4.2.2. *A necessary and sufficient condition for a function $f \in \mathbb{C}[0, \pi]$ to be positive definite on $\mathbb{S}^{d-1}$ is that f can be expressed in the form*

$$f(\theta) = \sum_{n=0}^{\infty} a_n C_n^{(\lambda)}(\cos(\theta))$$

where $\lambda = (d-2)/2$, $a_n \geq 0$, $\sum_{n=0}^{\infty} a_n C_n^{(\lambda)}(1) < \infty$.

Theorem 4.2.3. *A necessary condition for a function $f \in \mathbb{C}[0, \pi]$ to be positive definite on $\mathbb{S}^{\infty}$ is that f can be expressed in the form*

$$f(\theta) = \sum_{n=0}^{\infty} a_n \cos^n(\theta)$$

where $a_n \geq 0$, $\sum_{n=0}^{\infty} a_n < \infty$.

This necessary condition is in fact sufficient, as proved by Bingham [3].

5 Conditionally Positive Definite Functions

5.1 Prerequisites on finite dimensional spaces of functions

Unlike the case for positive definite functions, before defining conditionally positive definite functions, a few theorems will need to be established. To the best of my knowledge, the following treatment of conditionally positive definite functions is not found elsewhere.

Let Ω be a non-empty set and let U be a vector subspace of $\mathbb{R}^\Omega$, the set of all functions from Ω to $\mathbb{R}$. Recall Ω^*, the set of all finite linear combinations of point evaluation functional over $\mathbb{R}^\Omega$ (say δ_x is the point evaluation functional corresponding to $x \in \Omega$). Define $U^\perp$ to be the set of all functionals in Ω^* that annihilate U.That is to say that $\mu \in U^\perp$ if and only if both $\mu \in \Omega^*$ and $\mu u = 0$, $\forall u \in U$.

In turn, for any subspace Γ of Ω^*, let $\Gamma^\perp$ be the set of all functions in $\mathbb{R}^\Omega$ that are annihilated by all functionals in Γ. $\Gamma^\perp$ is clearly closed with respect to addition and scalar multiplication, and so is a subspace of $\mathbb{R}^\Omega$. U is **full** if $U^{\perp\perp} = U$. Let's start with some basic theorems concerning $\perp$.

Theorem 5.1.1. *Let Ω be a non-empty set and Let U and V be subspaces of $\mathbb{R}^\Omega$. The following hold*

(1) If $U \subseteq V$ then $V^\perp \subseteq U^\perp$.

(2) $\mathbb{R}_0^{\Omega\perp} = \Omega^*$, $\mathbb{R}^{\Omega\perp} = \Omega_0^*$.

(3) $\Omega^{*\perp} = \mathbb{R}_0^\Omega$, $\Omega_0^{*\perp} = \mathbb{R}^\Omega$.

Proof. For *(1)*, suppose $U \subseteq V$ and let $\mu \in V^\perp$. For any $u \in U$ it follows that $u \in V$ and so $\mu u = 0$. So $\mu \in U^\perp$, giving $V^\perp \subseteq U^\perp$.

For *(2)*, as the zero function is the only element of $\mathbb{R}_0^\Omega$, every functional in Ω^* annihilates $\mathbb{R}_0^\Omega$. So $\mathbb{R}_0^{\Omega\perp} = \Omega^*$. Let $\mu \in \mathbb{R}^{\Omega\perp}$, $\mu = \sum_{i=1}^n \lambda_i \delta_{x_i}$ for some $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \dots, \lambda_n) \in \mathbb{R}^n$ and $X = \{x_1, x_2, \dots, x_n\}$. if $\boldsymbol{\lambda} \neq \mathbf{0}$ then trivially a function f in $\mathbb{R}^\Omega$ can be found such that $\mu f \neq 0$. Hence μ is the zero functional.

For *(3)*, Let $f \in \Omega^{*\perp}$, as for any $x \in \Omega$, $\delta_x \in \Omega^*$ it follows that $\delta_x f = f(x) = 0$. Hence f is the zero function. Trivially every function in $\mathbb{R}^\Omega$ is annihilated by the zero functional, giving $\Omega_0^{*\perp} = \mathbb{R}^\Omega$. □

It follows from *Theorem* 5.1.1 that $\mathbb{R}^\Omega$ and $\mathbb{R}_0^\Omega$ are full. Here is an important theorem concerning full subspaces of $\mathbb{R}^\Omega$.

Theorem 5.1.2. *Let Ω be a non-empty set and let U be a subspace of $\mathbb{R}^\Omega$. If U is finite dimensional then U is full.*

Proof. This has already been established from *Theorem* 5.1.1 for $\mathbb{R}_0^\Omega$. Let U be a $m \in \mathbb{N}$ dimensional subspace of $\mathbb{R}^\Omega$. Proof by induction over the dimension of U.

Suppose $m = 1$. Let $f \in \mathbb{R}^\Omega$ such that f is annihilated by $U^\perp$. We want to show that $f \in U$, demonstrating that U is full. Let $u \in U$, u not the zero function. As U is of dimension 1, every element of U is a scalar multiple of u.

Let $x \in \Omega$ such that $u(x) \neq 0$ (which is guaranteed to exist as u is not the zero function). Let $y \in \Omega$. now consider the following functional in Ω^*

$$\mu = u(x)\delta_y - u(y)\delta_x$$

As $(u(x)\delta_y - u(y)\delta_x)u = u(x)u(y) - u(x)u(y) = 0$, μ annihilates U, that is to say $\mu \in U^\perp$. Recall f, which is annihilated by $U^\perp$, it follows that $\mu f = 0$. Suppose first that $f(x) = 0$, it follows that

$$0 = \mu f = u(x)f(y) - u(y)f(x) = u(x)f(y).$$

So $f(y) = 0$. As y is arbitrary, f must be the zero function, which is an element of U. Now suppose that $f(x) \neq 0$, then there is a non-zero $c \in \mathbb{R}$ such that $f(x) = cu(x)$. So

$$0 = \mu f = u(x)f(y) - u(y)f(x) = u(x)f(y) - cu(y)u(x).$$

So $f(y) = cu(y)$. As c is not dependent on y, it follows that $f = cu$, so $f \in U$. Hence U is full.

Now suppose for any subspace of dimension $k \in \mathbb{N}$ in $\mathbb{R}^\Omega$, said subspace is full. Suppose $m = k + 1$. Let $\{u_1, u_2, \ldots, u_k, u_{k+1}\}$ be a basis for U and let $f \in \mathbb{R}^\Omega$ such that f is not the zero function but is annihilated by $U^\perp$. The goal is to show once again that $f \in U$, demonstrating that U is full. Let V be the subspace of U spanned by $\{u_1, u_2, \ldots, u_k\}$, so the dimension of V is k. From our induction hypothesis, V is full.

If f is annihilated by $V^\perp$ then $f \in V \subset U$ as V is full. As such suppose f is not annihilated by $V^\perp$. Let $\mu \in V^\perp$ such that $\mu f \neq 0$. If $\mu u_{k+1} = 0$ then μ annihilates $\{u_1, u_2, \ldots, u_k, u_{k+1}\}$ and so annihilates U, implying that $\mu f = 0$ which contradicts our assumption. It follows that $\mu u_{k+1} \neq 0$. Let $\nu \in V^\perp$ and consider the following functional

$$(\mu u_{k+1})\nu - (\nu u_{k+1})\mu \in V^\perp.$$

This functional clearly annihilates u_{k+1} and given that it annihilates V, it follows that it annihilates U. So it must annihilate f, giving

$$(\mu u_{k+1})(\nu f) - (\nu u_{k+1})(\mu f) = 0.$$

Rearranging and letting $c = (\mu f)(\mu u_{k+1})^{-1}$ (recall that $\mu u_{k+1} \neq 0$), we obtain

$$\nu(f - cu_{k+1}) = 0.$$

Given that ν was an arbitrary element of $V^{\perp}$, $f - cu_{k+1} \in V$ as V is full. So $f \in U$. In all cases it has been shown that $f \in U$. Hence U is full.

It has been established that U is full for any $m \in \mathbb{N}$. $\square$

We will be concerned with the case when U is of finite dimension (say $m \in \mathbb{N}$).

Let $X \subseteq \Omega$. X is **unisolvent** with respect to U if the only element of U that is zero when evaluated at all points in X is the zero function.

Theorem 5.1.3. *Let Ω be a non-empty set and let U be a subspace of $\mathbb{R}^{\Omega}$ of dimension $m \in \mathbb{N}$. If $X \subseteq \Omega$ is unisolvent w.r.t U then there exists a finite subset of X that is unisolvent w.r.t U.*

Proof. Let $\{u_1, u_2, \ldots, u_m\}$ be a basis for U. Consider the set

$$S = \{(u_1(x), u_2(x), \ldots, u_m(x)) \mid x \in X\} \subseteq \mathbb{R}^m.$$

Take a maximal subset of S (say V) of linearly independent vectors in S, say $V = \{v_1, v_2, \ldots, v_d\}$ where d is clearly less than m. For each $v_i \in V$ there is a corresponding $y_i \in \Omega$ such that $v_i = (u_1(y_i), u_2(y_i), \ldots, u_m(y_i))$. Let $Y = \{y_1, y_2, \ldots, y_d\}$.

With this all set up, let $u \in U$ such that $u(y) = 0,\ \forall y \in Y$. Want to show that u must be the zero function, demonstrating that Y is a finite subset of Ω that is unisolvent w.r.t U. $u = \sum_{i=1}^{m} a_i u_i$ for some $a_i \in \mathbb{R}$. Let $x \in X$ and consider now the element of S $v_x = (u_1(x), u_2(x), \ldots, u_m(x))$ which is a linear combination of elements of V else $V \cup \{v_x\}$ would be a linearly independent set, contradicting the maximality of V. so $v_x = \sum_{i=1}^{d} b_i v_i$ for some $b_i \in \mathbb{R}$. This gives us the following: $u_k(x) = \sum_{i=1}^{d} b_i u_k(y_i)$. We then have the following

$$\begin{aligned} u(x) = \sum_{i=1}^{m} a_i u_i(x) = \sum_{i=1}^{m} a_i \sum_{j=1}^{d} b_j u_i(y_j) &= \sum_{j=1}^{d} b_j (\sum_{i=1}^{m} a_i u_i(y_j)) \\ &= \sum_{j=1}^{d} b_j u(y_j) = \sum_{j=1}^{d} b_j \cdot 0 = 0. \end{aligned}$$

Hence $u(x) = 0,\ \forall x \in X$. As X is unisolvent, u is the zero function. Hence Y is unisolvent w.r.t U. $\square$

There is an immediate corollary

Corollary 5.1.3.1. *Let Ω be a non-empty set and let U be a subspace of $\mathbb{R}^{\Omega}$ of dimension $m \in \mathbb{N}$. There exists a finite subset X of Ω such that X is unisolvent w.r.t U.*

Proof. Clearly Ω is unisolvent w.r.t U. By Theorem 5.1.3, there exists a finite subset of Ω that is unisolvent w.r.t U. $\square$

Theorem 5.1.4. *Let Ω be a non-empty set, U a $m \in \mathbb{N}$ dimensional subset of $\mathbb{R}^\Omega$ and $X = \{x_1, x_2, \dots, x_n\} \subseteq \Omega$ of size $n \in \mathbb{N}$. The following statements are equivalent:*

(1) X is unisolvent with respect to U.

(2) For any $u_1, u_2 \in U$, if $u_1(x_i) = u_2(x_i)$, $\forall i = 1, \dots, n$, then $u_1 = u_2$.

(3) Let $\{u_1, u_2, \dots, u_m\}$ be a basis of U. If $\sum_{i=1}^m c_i u_i(x_j) = 0$, $\forall j = 1, \dots, n$, where $c_i \in \mathbb{R}$, then $c_i = 0$, $\forall i = 1, \dots, m$.

(4) Let $B = \{u_1, u_2, \dots, u_m\}$ be a basis of U. Let $\boldsymbol{B}_X$ be the $n \times m$ matrix such that $(\boldsymbol{B}_X)_{ji} = u_i(x_j)$, $\forall i = 1, \dots, m$, $j = 1, \dots, n$. $\boldsymbol{B}_X$ is of full rank.

Proof. We will show first that *(1)* and *(2)* are equivalent.

Suppose *(2)*. $u \in U$ such that $u(x_i) = 0$, $\forall i = 1, \dots, n$. then $u(x_i) = 0(x_i)$, $\forall i = 1, \dots, n$ where 0 here is the zero fucntion of U. Hence $u = 0$, giving *(1)*.

Suppose *(1)*. Let $u_1, u_2 \in U$ such that $u_1(x_i) = u_2(x_i)$, $\forall i = 1, \dots, n$. It follows that $(u_1 - u_2)(x_i) = 0$, $\forall i = 1, \dots, n$. Hence $u_1 - u_2 = 0$ where 0 here is the zero functional giving *(2)*.

Now let's show that *(1)*, *(3)* and *(4)* are equivalent.

Suppose *(1)*. Let $\{u_1, u_2, \dots, u_m\}$ be a basis of U. Let $c_1, c_2, \dots, c_m \in \mathbb{R}$ such that $\sum_{i=1}^m c_i u_i(x_j) = 0$, $\forall j = 1, \dots, n$. By the unisolvency of X w.r.t U, $\sum_{i=1}^m c_i u_i$ is the zero function of U. Hence $c_i = 0$, $\forall i = 1, \dots, m$, giving *(3)*.

Suppose *(3)*. Let $\{u_1, u_2, \dots, u_m\}$ be a basis of U. Let $\boldsymbol{B}_X$ be the $n \times m$ matrix such that $(\boldsymbol{B}_X)_{ji} = u_i(x_j)$, $\forall i = 1, \dots, m$, $j = 1, \dots, n$. Let $\boldsymbol{c} = (c_1, c_2, \dots, c_m) \in \mathbb{R}^m$ such that $\boldsymbol{B}_X \boldsymbol{c} = \mathbf{0}$. By *(3)*, $\boldsymbol{c} = \mathbf{0}$ and so $\boldsymbol{B}_X$ is of full rank, giving *(4)*.

Finally, suppose *(4)*. Suppose $u(x_i) = 0$, $\forall i = 1, \dots, n$. $u = \sum_{i=1}^m c_i u_i$ where $\boldsymbol{c} = (c_1, c_2, \dots, c_m) \in \mathbb{R}^m$. It follows that $\boldsymbol{B}_X \boldsymbol{c} = \mathbf{0}$. As $\boldsymbol{B}_X$ is of full rank, $\boldsymbol{c} = \mathbf{0}$ and so $u = 0$, giving *(1)*. $\square$

Let $X = \{x_1, x_2, \dots, x_n\} \subseteq \Omega$ of $n \in \mathbb{N}$ distinct points of Ω. Let $U_X^\perp$ be the set of all $\boldsymbol{a} = (a_1, a_2, \dots, a_n) \in \mathbb{R}^n$ such that $\sum_{i=1}^n a_i \delta_{x_i} \in U^\perp$.

Theorem 5.1.5. *Let Ω be a non-empty set and let U be a subspace of $\mathbb{R}^\Omega$ of dimension $m \in \mathbb{N}$ with basis $B = \{u_1, u_2, \dots, u_m\}$ and $X = \{x_1, x_2, \dots, x_n\} \subseteq \Omega$ of size $n \in \mathbb{N}$ be unisolvent with respect to U. For any $\boldsymbol{a} = (a_1, a_2, \dots, a_n) \in \mathbb{R}^n$, $\boldsymbol{a} \in U_X^\perp$ if and only if $\boldsymbol{B}_X^T \boldsymbol{a} = 0$.*

Proof. This follows immediately from the fact that if $\boldsymbol{a} = (a_1, a_2, \dots, a_n) \in U_X^\perp$ then $\sum_{i=1} a_i u_k(x_i) = 0$, $\forall u_k \in B$. Expressing this in matrix form gives $\boldsymbol{B}_X^T \boldsymbol{a} = 0$. $\square$

Corollary 5.1.5.1. *Let Ω be a non-empty set and let U be a $m \in \mathbb{N}$ dimensional subspace of $\mathbb{R}^\Omega$. Let $X = \{x_1, x_2, \dots, x_n\} \subseteq \Omega$ of $n \in \mathbb{N}$ distinct points of Ω such that X is unisolvent w.r.t U. Then $n \geq m$ and $\dim U_X^\perp = n - m$.*

Proof. Let $B = \{u_1, u_2, \dots, u_m\}$ be a basis for U. By Theorem 5.1.4, the matrix $\boldsymbol{B}_X$ is off full rank m. By Theorem 5.1.5, $U_X^\perp = \{\boldsymbol{a} \in \mathbb{R}^n \mid \boldsymbol{B}_X^T \boldsymbol{a} = \boldsymbol{0}\}$. By the rank-nullity theorem $\operatorname{rank}(\boldsymbol{B}_X^T) + \operatorname{null}(\boldsymbol{B}_X^T) = n$. Rearranging gives $\operatorname{null}(\boldsymbol{B}_X^T) = n - m$. Clearly then $n \geq m$. $\square$

5.2 Introduction

In this section, a generalisation of positive definite functions is given and their basic general properties are studied. These conditionally positive definite functions, as the name implies, will be akin to positive definite functions but with constraints. More precisely, letting Ω be a non-empty set, a function $P : \Omega \times \Omega \to \mathbb{R}$ is **conditionally positive definite** over Ω with respect to U if

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j P(x_i, x_j) \geq 0, \tag{9}$$

for any pair $(X, \boldsymbol{a})$ where $X = \{x_1, x_2, \dots, x_n\} \subseteq \Omega$ of $n \in \mathbb{N}$ that is unisolvent w.r.t U and $\boldsymbol{a} = (a_1, a_2, \dots, a_n) \in U_X^\perp$. This is equivalent to the $n \times n$ matrix

$$\boldsymbol{P}_X := \begin{bmatrix} P(x_1, x_1) & P(x_1, x_2) & \dots & P(x_1, x_n) \\ P(x_2, x_1) & P(x_2, x_2) & \dots & P(x_2, x_n) \\ \vdots & \ddots & \ddots & \vdots \\ P(x_n, x_1) & P(x_n, x_2) & \dots & P(x_n, x_n) \end{bmatrix}$$

being non-negative definite with respect to $U_X^\perp$. That is to say that for any $\boldsymbol{a} = (a_1, a_2, \dots, a_n) \in U_X^\perp$

$$\boldsymbol{a}^T \boldsymbol{P}_X \boldsymbol{a} \geq 0.$$

The above matrix $\boldsymbol{P}_X$ is the **Gramian matrix** of P at X. The common setting to consider such functions is $\Omega = \mathbb{R}^d$, $d \in \mathbb{N}$ and U is usually taken to be a set of low degree polynomials. P is **strictly conditionally positive definite** (**SCPD**) if the inequality in (9) is strict when $\boldsymbol{a} \neq \boldsymbol{0}$.

Again our concern will be with conditionally positive definite functions that are symmetric in the sense that

$$P(x, y) = P(y, x), \ \forall x, y \in \Omega.$$

Recall Ω^* be the set of all finite linear combinations of point evaluation functionals over $\mathbb{R}^\Omega$. Another equivalent characterisation of conditionally positive definite functions can be given. A function $P : \Omega \times \Omega \to \mathbb{R}$ is conditionally positive definite with respect to U if for any $\mu \in U^\perp$

$$\mu^{(1)} \mu^{(2)} P \geq 0.$$

From this, it should be clear that a SCPD function P over Ω w.r.t U gives rise to semi inner products $\langle\cdot,\cdot\rangle$ over $U^\perp$ via

$$\langle\mu_1,\mu_2\rangle = \mu_1^{(1)}\mu_2^{(2)}P.$$

When P is symmetric, $\mu_1^{(1)}\mu_2^{(2)}P = \mu_2^{(1)}\mu_1^{(2)}P$, As such the superscripts will be left out so $\mu_1^{(1)}\mu_2^{(2)}P = \mu_1\mu_2 P$.

The following theorem is analogous to Theorem 2.1.1.

Theorem 5.2.1. *Let Ω be a non-empty set and let U be a $m \in \mathbb{N}$ dimensional space of functions from $\Omega \to \mathbb{R}$. Let P be a symmetric CPD function over Ω w.r.t U. Let $X = \{x_1, x_2, \ldots, x_n\}, Y = \{y_1, y_2, \ldots, y_{n'}\} \subseteq \Omega$ of size $n, m \in \mathbb{N}$ respectively be unisolvent w.r.t U and $\boldsymbol{a} = (a_1, a_2, \ldots, a_n) \in U_X^\perp$, $\boldsymbol{b} = (b_1, b_2, \ldots, b_m) \in U_Y^\perp$. The following holds*

$$|\sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_j P(x_i, y_j)|^2 \leq \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j P(x_i, x_j) \sum_{k=1}^{m}\sum_{l=1}^{m} b_k b_l P(y_k, y_l).$$

Proof. Let $Z = X \cup Y = \{z_1, z_2, \ldots, z_N\}$ As X and Y may not be disjoint, it isn't necessarily the case that $N = n + n'$. It should be clear that Z is unisolvent with respect to U given both X and Y are. We wish to find a $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_N), \boldsymbol{\gamma} = (\gamma_1, \gamma_2, \ldots, \gamma_N) \in U_Z^\perp$ such that

$$\begin{aligned}
\boldsymbol{\lambda}^T \boldsymbol{P}_Z \boldsymbol{\gamma} &= \sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_j P(x_i, y_j),\\
\boldsymbol{\lambda}^T \boldsymbol{P}_Z \boldsymbol{\lambda} &= \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j P(x_i, x_j),\\
\boldsymbol{\gamma}^T \boldsymbol{P}_Z \boldsymbol{\gamma} &= \sum_{i=1}^{m}\sum_{j=1}^{m} b_i b_j P(y_i, y_j).
\end{aligned}$$

Let $\boldsymbol{Q}$ be an $N \times m$ matrix of full rank whose range is $U_Z^\perp$ It follows that the $m \times m$ matrix $\boldsymbol{Q}^T \boldsymbol{P}_Z \boldsymbol{Q}$ is symmetric non negative definite as

$$\boldsymbol{v}^T(\boldsymbol{Q}^T \boldsymbol{P}_Z \boldsymbol{Q})\boldsymbol{v} = (\boldsymbol{Q}\boldsymbol{v})^T \boldsymbol{P}_Z (\boldsymbol{Q}\boldsymbol{v}) \geq 0,$$

for any $\boldsymbol{v} \in \mathbb{R}^m$ as the matrix $\boldsymbol{P}_Z$ is non-negative definite over $U_Z^\perp$. Symmetry follows from the symmetry of $\boldsymbol{P}_Z$.

Let $\boldsymbol{\lambda}', \boldsymbol{\gamma}' \in \mathbb{R}^m$ such that $\boldsymbol{Q}\boldsymbol{\lambda}' = \boldsymbol{\lambda}$ and $\boldsymbol{Q}\boldsymbol{\gamma}' = \boldsymbol{\gamma}$. It follows from the symmetric non-negative definiteness of $\boldsymbol{Q}^T \boldsymbol{P}_Z \boldsymbol{Q}$ that,

$$(\boldsymbol{\lambda}'^T \boldsymbol{Q}^T \boldsymbol{P}_Z \boldsymbol{Q} \boldsymbol{\gamma}')^2 \leq (\boldsymbol{\lambda}'^T \boldsymbol{Q}^T \boldsymbol{P}_Z \boldsymbol{Q} \boldsymbol{\lambda}')(\boldsymbol{\gamma}'^T \boldsymbol{Q}^T \boldsymbol{P}_Z \boldsymbol{Q} \boldsymbol{\gamma}').$$

Let's note that the above inequality is the inequality that we desire to prove. The appropriate choices for $\boldsymbol{\lambda}$ and $\boldsymbol{\gamma}$ are:

$$\boldsymbol{\lambda}_i = \begin{cases} a_k & \text{if } z_i = x_k, \quad \text{for some } k = 1, \dots, n \\ 0 & \text{otherwise} \end{cases},$$

$$\boldsymbol{\gamma}_i = \begin{cases} b_k & \text{if } z_i = y_k, \quad \text{for some } k = 1, \dots, m \\ 0 & \text{otherwise} \end{cases}.$$

This completes the proof. □

There is an immediate corollary to this theorem

Corollary 5.2.1.1. *Let Ω be a non-empty set and let U be a $m \in \mathbb{N}$ dimensional space of functions from $\Omega \to \mathbb{R}$. Let P be a symmetric CPD function over Ω w.r.t U and let $\mu \in U^{\perp}$. If $\mu\mu P = 0$ then $\mu P \in U$.*

Proof. Let $\nu \in U^{\perp}$. As $\mu\mu P = 0$, by Theorem 5.2.1

$$|\nu\mu P|^2 \leq \nu\nu P\mu\mu P = \nu\nu P \cdot 0 = 0.$$

Hence $\nu\mu P = 0$. This implies that $U^{\perp}$ annihilates the function $\mu P \in \mathbb{R}^{\Omega}$. As U is full, $\mu P \in U$. □

5.3 Linear Interpolation with a Kernel Revisited

Let Ω be a non-empty set and $F : \Omega \times \Omega \to \mathbb{R}$. Suppose we wish to find a function $s : \Omega \to \mathbb{R}$ that satisfy the following interpolation constraints

$$s(x_i) = d_i, \ \forall i = 1, \dots, n,$$

where $X = \{x_1, x_2, \dots, x_n\} \subseteq \Omega$ of size $n \in \mathbb{N}$ and $\boldsymbol{d} = (d_1, d_2, \dots, d_n) \in \mathbb{R}^n$. Unlike in Section 2.2 however, There is a twist: Let U be a vector space of dimension $m \in \mathbb{N}$ of functions from Ω to $\mathbb{R}$. If there exists a function $u \in U$ such that $u(x_i) = d_i, \ \forall i = 1, \dots, n$, then we have found such an s, as desired. It would be preferable if u were the only choice in U that satisfies the interpolation constraints, this is in fact equivalent to X being unisolvent with respect to U, which will be shown in the next subsection.

More generally, possibly such an s can be found in the form

$$s(\,\cdot\,) = \sum_{i=1}^{n} \lambda_i F(x_i, \,\cdot\,) + u \tag{10}$$

where $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \dots, \lambda_n) \in \mathbb{R}^n$ and $u \in U$. It would follow that

$$\sum_{i=1}^{n} \lambda_i F(x_i, x_k) + u(x_k) = d_k, \ \forall k = 1, \dots, n.$$

Letting $B = \{u_1, u_2, \dots, u_m\}$ be a basis for U and letting $\boldsymbol{c} = (c_1, c_2, \dots, c_m) \in \mathbb{R}^m$ such that $u = \sum_{i=1}^m c_i u_i$ we obtain

$$\sum_{i=1}^{n} \lambda_i F(x_i, x_k) + \sum_{i=1}^{m} c_i u_i(x_k) = d_k, \ \forall k = 1, \dots, n.$$

or equivalently, defining the following matrices

$$\boldsymbol{F}_X := \begin{bmatrix} F(x_1, x_1) & F(x_1, x_2) & \dots & F(x_1, x_n) \\ F(x_2, x_1) & F(x_2, x_2) & \dots & F(x_2, x_n) \\ \vdots & \ddots & \ddots & \vdots \\ F(x_n, x_1) & F(x_n, x_2) & \dots & F(x_n, x_n) \end{bmatrix}$$

$$\boldsymbol{B}_X := \begin{bmatrix} u_1(x_1) & u_2(x_1) & \dots & u_m(x_1) \\ u_1(x_2) & u_2(x_2) & \dots & u_m(x_2) \\ \vdots & \ddots & \ddots & \vdots \\ u_1(x_n) & u_2(x_n) & \dots & u_m(x_n) \end{bmatrix}$$

it follows that

$$\begin{bmatrix} \boldsymbol{F}_X & \boldsymbol{B}_X \end{bmatrix} \begin{bmatrix} \boldsymbol{\lambda} \\ \boldsymbol{c} \end{bmatrix} = \boldsymbol{d}.$$

As this matrix is under-determined, There is no unique interpolant of this form unless additional constraints are added. Consider the additional constraint $\boldsymbol{B}_X^T \boldsymbol{\lambda} = \boldsymbol{0}$. It will be shown in the next subsection that this constraint is equivalent to $\boldsymbol{\lambda} \in U_X^{\perp}$. Adding this constraint gives the black matrix equation

$$\begin{bmatrix} \boldsymbol{F}_X & \boldsymbol{B}_X \\ \boldsymbol{B}_X^T & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{\lambda} \\ \boldsymbol{c} \end{bmatrix} = \begin{bmatrix} \boldsymbol{d} \\ \boldsymbol{0} \end{bmatrix}.$$

If F is conditionally positive definite over Ω w.r.t U, then this block matrix is invertible.

Theorem 5.3.1. *Let Ω be a non-empty set and let U be a subspace of $\mathbb{R}^\Omega$ of dimension $m \in \mathbb{N}$ with basis $B = \{u_1, u_2, \dots, u_m\}$. Let P be a SCPD function over Ω w.r.t U and $X = \{x_1, x_2, \dots, x_n\} \subseteq \Omega$ of size $n \in \mathbb{N}$ that is unisolvent w.r.t U. The block matrix*

$$\begin{bmatrix} \boldsymbol{P}_X & \boldsymbol{B}_X \\ \boldsymbol{B}_X^T & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{\lambda} \\ \boldsymbol{c} \end{bmatrix} = \boldsymbol{0}$$

has only the trivial solution for the pair $(\boldsymbol{\lambda}, \boldsymbol{c})$.

Proof. Left multiplying by the transpose of the vector containing $(\boldsymbol{\lambda}, \boldsymbol{c})$ gives

$$\boldsymbol{\lambda}^T \boldsymbol{P}_X \boldsymbol{\lambda} + 2\boldsymbol{\lambda}^T \boldsymbol{B}_X \boldsymbol{c} = \boldsymbol{\lambda}^T \boldsymbol{P}_X \boldsymbol{\lambda} = \boldsymbol{0}.$$

As P is SCPD over Ω, $\boldsymbol{\lambda} = \boldsymbol{0}$. So the above block matrix system reduces down to $\boldsymbol{B}_X \boldsymbol{c} = \boldsymbol{0}$. As *boldsymbolB*$_X$ is of full rank, by Theorem 5.1.4, $\boldsymbol{c} = \boldsymbol{0}$. Hence there is only a trivial solution. □

6 Semi Reproducing Kernel Hilbert Spaces

6.1 Introduction

This introduction section is based heavily on Chapter 2 of the Variational theory of splines (Bezhaev & Valisenko [2]). Keep in mind however the terminology used here will be different. See Appendix A for basic theory of semi inner products. The notion of a reproducing kernel Hilbert space can be extended from what we have considered so far in Chapter 3.

Let Ω be a non-empty set and let $(H, \langle \cdot, \cdot \rangle)$ be a real semi Hilbert space with null space U of dimension $m \in \mathbb{N}$. Introduce a real inner product $[\cdot, \cdot]$ over H such that $(H, [\cdot, \cdot])$ is a RKHS over Ω. $\langle \cdot, \cdot \rangle$ and $[\cdot, \cdot]$ will be related as follows: there exists a $C \in \mathbb{R}$ such that

$$\|f\| \leq C|f|, \ \ \forall f \in H,$$

where $\| \cdot \|$ is the corresponding semi norm of $\langle \cdot, \cdot \rangle$ and $| \cdot |$ is the corresponding norm of $[\cdot, \cdot]$. The space $(H, \langle \cdot, \cdot \rangle, [\cdot, \cdot])$ is a **Semi Reproducing Kernel Hilbert Space** (**SRKHS**). H^* will refer to the set of all bounded linear functionals with respect to the inner product $[\cdot, \cdot]$. Let $U^{\perp}$ be the subspace of H^* of all continuous linear functionals that annihilate U.

As $(H, [\cdot, \cdot])$ is a RKHS, all point evaluation functionals over H are continuous. It follows that if a given finite linear combination of point evaluations annihilates U, then it resides in $U^{\perp}$. Note that this definition of $U^{\perp}$ differs from that of the previous section but still contains all finite linear combination of point evaluations that annihilates U. In this section.

A function $K : \Omega \times \Omega \to \mathbb{R}$ is a semi reproducing kernel of $(H, \langle \cdot, \cdot \rangle, [\cdot, \cdot])$ if

1. $\mu K \in H$, $\forall \mu \in H^*$,
2. If $\mu \in U^{\perp}$, then $\mu f = \langle f, \mu K \rangle$, $\forall f \in H$.

We have the following theorem from [2].

Theorem 6.1.1. *Let Ω be a non-empty set and let $(H, \langle \cdot, \cdot \rangle, [\cdot, \cdot])$ be SRKHS. There exists a symmetric semi repoducing kernel K of $(H, \langle \cdot, \cdot \rangle, [\cdot, \cdot])$. It is not unique.*

The following theorem should be of no surprise.

Theorem 6.1.2. *Let Ω be a non-empty set and let $(H, \langle \cdot, \cdot \rangle, [\cdot, \cdot])$ be SRKHS with symmetric semi reproducing kernel K. K is symmetric conditionally positive definite over Ω w.r.t U.*

Proof. Symmetry is trivial. Let $X = \{x_1, x_2, \dots, x_n\} \subseteq \Omega$ of size $n \in \mathbb{N}$ and let $\boldsymbol{a} = (a_1, a_2, \dots, a_n) \in U_X^{\perp}$. The linear functional

$$\sum_{i=1}^{n} a_i \delta_{x_i}$$

is in H^* as $(H, \langle\cdot,\cdot\rangle)$ is a RKHS. It also annihilates U and so is in $U^{\perp}$. As K is a semi reproducing kernel, it follows that $\sum_{i=1}^{n} a_i K_{x_i} \in H$ and

$$0 \leq \langle \sum_{i=1}^{n} a_i K_{x_i}, \sum_{i=1}^{n} a_i K_{x_i} \rangle = \sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j K(x_i, x_j).$$

Hence K is conditionally positive definite. □

6.2 Interpolation and Smoothing in Semi Reproducing Kernel Hilbert Spaces

6.2.1 Interpolation with point evaluations for Semi Reproducing Kernels

The result concerning the linear systems to solve to find these interpolation and approximations are also known (See [9], [15]). However, to the best of my knowledge the proofs given here for the energy penalized least squares conditions are novel.

For our purposes, this notion of a semi reproducing kernel Hilbert space will be generalised. For convenience, to say that $(\Omega, H, \langle\cdot,\cdot\rangle, \|\cdot\|, U, K)$ is a SRKHS is to say that $(H, \langle\cdot,\cdot\rangle)$ is a SRKHS over Ω, a non empty set, with null space U and reproducing kernel K. $\|\cdot\|$ being the natural norm induced by $\langle\cdot,\cdot\rangle$. K is a reproducing kernel of H in the following sense: For any choice of $X = \{x_1, x_2, \ldots, x_n\} \subseteq \Omega$ of some size $n \in \mathbb{N}$ that is unisolvent with respect to U, the following holds

(1) $\sum_{i=1}^{n} \lambda_i K_{x_i} \in H, \;\; \forall \boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_n) \in U_X^{\perp}$.

(2) $\langle f, \sum_{i=1}^{n} \lambda_i K_{x_i} \rangle = \sum_{i=1}^{n} \lambda_i f(x_i) \;\; \forall \boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_n) \in U_X^{\perp}$.

Clearly K is conditionally positive definite w.r.t U. That is to say, for any $X = \{x_1, x_2, \ldots, x_n\} \subseteq \Omega$ that is unisolvent with respect to U and $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_n) \in U_X^{\perp}$, the following holds

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \lambda_j K(x_i, x_j) \geq 0.$$

Furthermore, K is strictly conditionally positive definite w.r.t U if the inequality above is strict whenever $\boldsymbol{\lambda} \neq \mathbf{0}$. It should be clear that all RKHS's are SRKHS's with null space $U = \{0\}$, in which case, any finite subset $X = \{x_1, x_2, \ldots, x_n\} \subseteq \Omega$ is unisolvent and $U_X^{\perp} = \mathbb{R}^n$. Let $\boldsymbol{K}_X$ be the $n \times n$ matrix with entries $(\boldsymbol{K}_X)_{ij} = K(x_i, x_j)$. Recall that this matrix is the Gramian matrix of K at X. Clearly the matrix $\boldsymbol{K}_X$ is conditionally positive definite in the sense that

$$\boldsymbol{\lambda}^T \boldsymbol{K}_X \boldsymbol{\lambda} \geq 0, \; \forall \boldsymbol{\lambda} \in U_X^{\perp},$$

with the equality being strict for non-zero $\boldsymbol{\lambda}$ when K is strictly conditionally positive definite. H is **separating** if for any $X = \{x_1, x_2, \dots, x_n\} \subseteq \Omega$ unisolvent w.r.t U and for any $\boldsymbol{d} = (d_1, d_2, \dots, d_n) \in \mathbb{R}^n$ there exists an $f \in H$ such that $f(x_i) = d_i,\ \forall i = 1, \dots, n$. A SRKHS $(\Omega, H, \langle\cdot,\cdot\rangle, \|\cdot\|, U, K)$ is strict if H is separating and K is strictly conditionally positive definite w.r.t U.

Theorem 6.2.1. *Consider the following minimisation problem.*

Let $(\Omega, H, \langle\cdot,\cdot\rangle, \|\cdot\|, U, K)$ be a strict SRKHS. Let $X = \{x_1, x_2, \dots, x_n\} \subseteq \Omega$, X unisolvent w.r.t U. Find $s \in H$ such that

$$s = \underset{f \in H}{\arg\min} \|f\|^2 \text{ subject to } f(x_i) = d_i, \ \ \forall i = 1 \dots n \tag{11}$$

where $\boldsymbol{d} = (d_1, d_2, \dots, d_n) \in \mathbb{R}^n$. The solution to this problem is unique and given by

$$s(\,\cdot\,) = \sum_{i=1}^{n} \lambda_i K_{x_i}(\,\cdot\,) + u^*(\,\cdot\,)$$

where $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \dots, \lambda_n) \in U_X^{\perp}$ and $u^* \in U$. Letting $B = \{u_1, u_2, \dots, u_m\}$ be an arbitrary basis for U and $\boldsymbol{c} = (c_1, c_2, \dots, c_m) \in \mathbb{R}^m$ such that $u* = \sum_{i=1}^{m} c_i u_i$, we have the following linear system:

$$\begin{bmatrix} \boldsymbol{K}_X & \boldsymbol{B}_X \\ \boldsymbol{B}_X^T & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{\lambda} \\ \boldsymbol{c} \end{bmatrix} = \begin{bmatrix} \boldsymbol{d} \\ \boldsymbol{0} \end{bmatrix}$$

Furthermore, this linear system has a unique solution for $(\boldsymbol{\lambda}, \boldsymbol{c})$, the parameters of the smoothest interpolant.

Proof. The above block matrix system indeed has a unique solution by Theorem 5.3.1. By the unisolvency of X w.r.t U, $\boldsymbol{B}_X^T$ has linearly independent columns. Hence the second row of the block system implies that the functional $\sum_{i=1}^{n} \lambda_i \delta_{x_i}$ annihilates U. So the function $\sum_{i=1}^{n} \lambda_i K_{x_i}$ is in H. This immediately implies that $s \in H$. We proceed to show that

$$s(\,\cdot\,) = \sum_{i=1}^{n} \lambda_i K_{x_i}(\,\cdot\,) + \sum_{j=1}^{m} c_j u_j(\,\cdot\,)$$

is a solution to the minimisation problem. First it satisfies the interpolation constraints as

$$s(x_k) = \sum_{i=1}^{n} \lambda_i K_{x_i}(x_k) + \sum_{j=1}^{m} c_j u_j(x_k) = (\boldsymbol{K}_X)_k \boldsymbol{\lambda} + (\boldsymbol{B}_X)_k \boldsymbol{c} = d_k.$$

where $(\boldsymbol{K}_X)_k$ and $(\boldsymbol{B}_X)_k$ are the k^{th} row vector of $\boldsymbol{K}_X$ and $\boldsymbol{B}_X$ respectively. To show that s minimises the norm, let $s' \in H$ such that $s'(x_i) = d_i, \ \ \forall i = 1 \dots n$.

As

$$\begin{aligned}
\langle s' - s, s\rangle &= \langle s' - s, \sum_{i=1}^{n} \lambda_i K_{x_i} + \sum_{j=1}^{m} c_j u_j\rangle \\
&= \langle s' - s, \sum_{i=1}^{n} \lambda_i K_{x_i}\rangle \\
&= \sum_{i=1}^{n} \lambda_i (s' - s)(x_i) = \sum_{i=1}^{n} \lambda_i (d_i - d_i) = 0,
\end{aligned}$$

the following holds

$$\|s'\|^2 = \|s' - s + s\|^2 = \|s' - s\|^2 + \|s\|^2 - 2\langle s' - s, s\rangle = \|s\|^2 + \|s - s'\|^2.$$

$\|s - s'\|^2 \geq 0$, so $\|s'\|^2 \geq \|s\|^2$. This shows that s minimises the norm subject to the given constraints. For uniqueness, Suppose $\|s'\| = \|s\|$. Rearranging the equation above gives $\|s - s'\|^2 = \|s'\|^2 - \|s\|^2$. As $\|s'\|^2 - \|s\|^2 = 0$, $s' = s$. □

From this theorem, it is natural to define the following subspace. Let $\mathcal{K}_X$ be the **approximation subspace** of H w.r.t unisolvent set X. That is to say that

$$\mathcal{K}_X = \{\sum_{i=1}^{n} \lambda_i K_{x_i}(\,\cdot\,) + u \in H \mid \boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_n) \in U_X^{\perp},\ u \in U\}.$$

Often $\mathcal{K}_X$ will simply be referred to as $\mathcal{K}$. The interpolation space $\mathcal{K}$ has a dimension of n given the additional assumption that $\sum_{i=1}^{n} \lambda_i K_{x_i} \notin U$, $\forall \boldsymbol{\lambda} \in U_X^{\perp}$ unless $\boldsymbol{\lambda} = \mathbf{0}$. This is because the dimension of $U_X^{\perp}$ is $n - m$ and the dimension of U is m. The interpolant of the above theorem is a natural projection $\mathcal{I}_X : H \to \mathcal{K}_X$ such that

$$(\mathcal{I}_X f)(x_i) = f(x_i),\ \forall f \in H,$$

The above theorem shows that this function is well defined and has a natural orthogonal projection $I - \mathcal{I}_X$. To see this, first let $\mathcal{I}_X f = \sum_{i=1}^{n} \lambda_i K_{x_i} + u^*$ where $(\lambda_1, \lambda_2, \ldots, \lambda_n) \in U_X^{\perp}$ and $u \in U$.

$$\begin{aligned}
\langle \mathcal{I}_X f, (I - \mathcal{I}_X) f\rangle &= \langle \sum_{i=1}^{n} \lambda_i K_{x_i} + u^*, (I - \mathcal{I}_X) f\rangle \\
&= \langle \sum_{i=1}^{n} \lambda_i K_{x_i}, (I - \mathcal{I}_X) f\rangle \\
&= \sum_{i=1}^{n} \lambda_i ((I - \mathcal{I}_X) f)(x_i) \\
&= \sum_{i=1}^{n} \lambda_i (f(x_i) - f(x_i)) = 0.
\end{aligned}$$

For any given $f \in H$, $\mathcal{I}_X f$ will be referred to as the interpolant of f over X (or just the interpolant of f).

6.2.2 Penalised Least Squares for Semi Reproducing Kernels

Not let's consider the following penalised least squares problem. Recall that we are working in a strict SRKHS $(\Omega, H, \langle\cdot,\cdot\rangle, \|\cdot\|, U, K)$. Let $\boldsymbol{w} = (w_1, w_2, \ldots, w_n) \in \mathbb{R}^n$ be a vector of positive weights, $X = \{x_1, x_2, \ldots, x_n\} \subseteq \Omega$ be unisolvent w.r.t U. $\mathcal{K} := \mathcal{K}_X$ is the approximation subspace of H with respect to X and $\mathcal{I}_X$ is the natural projection onto $\mathcal{K}$. Let the semi inner product $[\cdot,\cdot]_X^{\boldsymbol{w}}$ over H be defined as follows:

$$[f, g]_X^{\boldsymbol{w}} = \sum_{i=1}^{n} w_i f(x_i) g(x_i) \quad \forall f, g \in H.$$

From the unisolvency of X w.r.t U, this semi inner product is a strict inner product over U. Let $\boldsymbol{d} = (d_1, d_2, \ldots, d_n) \in \mathbb{R}^n$ and let $s \in H$ such that $s(x_i) = d_i,\ \forall i = 1, \ldots, n$ (such an s is guaranteed to exist due to H being separating). Let $E_X^{\boldsymbol{w},\boldsymbol{d}} : H \to \mathbb{R}$ be defined as follows:

$$\begin{aligned} E_X^{\boldsymbol{w},\boldsymbol{d}}(f) &= \langle f, f\rangle + [f - s, f - s]_X^{\boldsymbol{w}} \\ &= \langle f, f\rangle + \sum_{i=1}^{n} w_i (f(x_i) - d_i)^2, \quad \forall f \in H. \end{aligned}$$

The penalised least squares problem is to find a function in H that minimises $E_X^{\boldsymbol{w},\boldsymbol{d}}$. It will be shown that there is a unique minimiser of this function in H. This will be proven in three stages: first the existence of a unique minimiser will be shown for when the energy is restricted to be defined over U. Second it will be shown when restricted to any affine subset of the form $f + U,\ f \in \mathcal{K}$. Finally, it will be established for the whole of H.

Let $F_X^{\boldsymbol{w},\boldsymbol{d}} : U \to \mathbb{R}$ be defined as follows:

$$F_X^{\boldsymbol{w},\boldsymbol{d}}(u) = [u - s, u - s]_X^{\boldsymbol{w}} = \sum_{i=1}^{n} w_i (u(x_i) - d_i)^2.$$

It should be clear that $F_X^{\boldsymbol{w},\boldsymbol{d}}$ is $E_X^{\boldsymbol{w},\boldsymbol{d}}$ restricted to U.

Lemma 6.2.2. *Let $(\Omega, H, \langle\cdot,\cdot\rangle, \|\cdot\|, U, K)$ be a strict SRKHS and let $X = \{x_1, x_2, \ldots, x_n\} \subseteq \Omega$ be unisolvent w.r.t U. $F_X^{\boldsymbol{w},\boldsymbol{d}}$, defined as above, has a unique minimiser over U.*

Proof. Let $B = \{u_1, u_2, \ldots, u_m\}$ be an basis for U. We can transform $F_X^{\boldsymbol{w},\boldsymbol{d}}$ into a function over $\mathbb{R}^m$. Let $u \in U$, $u = \sum_{i=1}^{m} c_i u_i$ where $\boldsymbol{c} = (c_1, c_2, \ldots, c_m) \in \mathbb{R}^m$.

$$\begin{aligned} F^{\boldsymbol{w},\boldsymbol{d}}(u) = \sum_{i=1}^{n} w_i (u(x_i) - d_i)^2 &= \sum_{i=1}^{n} w_i (\sum_{j=1}^{m} c_j u_j(x_i) - d_i)^2 \\ &= (\boldsymbol{B}_X \boldsymbol{c} - \boldsymbol{d})^T \boldsymbol{W} (\boldsymbol{B}_X \boldsymbol{c} - \boldsymbol{d}), \end{aligned}$$

where $\boldsymbol{W}$ is the $n \times n$ diagonal matrix with diagonal entries $(\boldsymbol{W})_{ii} = w_i \ \ \forall i = 1, \ldots, n$. Let $\mathcal{F}_X^{\boldsymbol{w},\boldsymbol{d}} : \mathbb{R}^m \to \mathbb{R}$ be defined as follows

$$\mathcal{F}_X^{\boldsymbol{w},\boldsymbol{d}}(\boldsymbol{c'}) = (\boldsymbol{B}_X\boldsymbol{c'} - \boldsymbol{d})^T \boldsymbol{W}(\boldsymbol{B}_X\boldsymbol{c'} - \boldsymbol{d}),\ \forall \boldsymbol{c'} \in \mathbb{R}^m.$$

Consider now the inner product over $\mathbb{R}^n$ $\langle \cdot, \cdot \rangle_{\boldsymbol{W}} : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ be $\langle \boldsymbol{v}, \boldsymbol{u} \rangle_{\boldsymbol{W}} = \boldsymbol{v}^T \boldsymbol{W} \boldsymbol{u}$ and let $\| \cdot \|_{\boldsymbol{W}}$ be the associated norm.

$$\mathcal{F}_X^{\boldsymbol{w},\boldsymbol{d}}(\boldsymbol{c'}) = \|\boldsymbol{B}_X\boldsymbol{c'} - \boldsymbol{d}\|_{\boldsymbol{W}},\ \forall \boldsymbol{c'} \in \mathbb{R}^m.$$

Let $V \subseteq \mathbb{R}^n$ be the range of $\boldsymbol{B}_X$. The theory of inner product spaces establishes that there exists a unique $\boldsymbol{v^*} \in V$ of minimal distance to $\boldsymbol{d}$ characterised by $\boldsymbol{v^*} - \boldsymbol{d} \perp V$. As X is unisolvent w.r.t U, $\boldsymbol{B}_X$ is of full rank and so there exists a unique $\boldsymbol{c^*} = (c_1^*, c_2^*, \ldots, c_m^*) \in \mathbb{R}^m$ such that $\boldsymbol{B}_X\boldsymbol{c^*} = \boldsymbol{v^*}$. The corresponding function $u^* = \sum_{i=1}^m c_i^* u_i$ is then the unique minimiser of $F_X^{\boldsymbol{w},\boldsymbol{d}}$ over U. □

Theorem 6.2.3. *Let* $(\Omega, H, \langle\cdot,\cdot\rangle, \|\cdot\|, U, K)$ *be a strict SRKHS and* $X = \{x_1, x_2, \ldots, x_n\} \subseteq \Omega$ *be unisolvent w.r.t U. Let* $f + U$ *be an affine subset of the approximation subspace* $\mathcal{K}_X$. $E_X^{\boldsymbol{w},\boldsymbol{d}}$*, defined as above, has a unique minimiser over* $f + U$.

Proof. let $f + u \in f + U$. $E_X^{\boldsymbol{w},\boldsymbol{d}}(f + u) = \langle f, f \rangle + \sum_{i=1}^n w_i(f(x_i) + u(x_i) - d_i)^2$. Letting $\boldsymbol{e} = (f(x_1) - d_1, f(x_2) - d_2, \ldots, f(x_n) - d_n) \in \mathbb{R}^n$, we see that $E_X^{\boldsymbol{w},\boldsymbol{d}}(f+u) = \langle f, f \rangle + E_X^{\boldsymbol{w},\boldsymbol{e}}(u)$. Minimising $E_X^{\boldsymbol{w},\boldsymbol{d}}$ over $f + U$ is then equivalent to minimising $E_X^{\boldsymbol{w},\boldsymbol{e}}$ over U. This is equivalent to minimising $F_X^{\boldsymbol{w},\boldsymbol{e}}$ over U. From Lemma 6.2.2. a unique minimiser $u^* \in U$ exists for $F_X^{\boldsymbol{w},\boldsymbol{e}}$. So $f + u^*$ is the unique minimiser of $E_X^{\boldsymbol{w},\boldsymbol{d}}$. □

Consider the quotient space $\mathcal{K}/U$, that is, the set of all affine subsets of the form $f + U \subset \mathcal{K}$ (which is equivalent to saying $f \in \mathcal{K}$). Let $(\cdot, \cdot)$ be the natural inner product over $\mathcal{K}/U$, so

$$(f + U, g + U) = \langle f, g \rangle,\ \forall f, g \in \mathcal{K}.$$

This is well defined, as established in Section A. From Theorem 6.2.3, the energy $E_X^{\boldsymbol{w},\boldsymbol{d}}$ has a unique minimiser over every affine subset $f + U \in \mathcal{K}/U$. This allows us to give every affine subset $f + U \in \mathcal{K}/U$ a natural energy, namely, the energy of the unique minimiser of $E_X^{\boldsymbol{w},\boldsymbol{d}}$ over $f + U$. The natural energy $\widehat{E}_X^{\boldsymbol{w},\boldsymbol{d}} : \mathcal{K}/U \to \mathbb{R}$ is then defined as follows:

$$\widehat{E}_X^{\boldsymbol{w},\boldsymbol{d}}(f + U) = \min_{g \in f+U} E_X^{\boldsymbol{w},\boldsymbol{d}}(g).$$

It will be shown that this function is continuous and strictly convex over $\mathcal{K}/U$. Noting that

$$(f + U, f + U) = \langle g, g \rangle,\ \forall f \in \mathcal{K}, g \in f + U,$$

$$\widehat{E}_X^{\boldsymbol{w},\boldsymbol{d}}(f+U) = \min_{g\in f+U} E_X^{\boldsymbol{w},\boldsymbol{d}}(g) = \min_{g\in f+U}(\langle g,g\rangle + \sum_{i=1}^n w_i(g(x_i)-d_i)^2)$$
$$= (f+U,f+U) + \min_{g\in f+U}(\sum_{i=1}^n w_i(g(x_i)-d_i)^2).$$

Theorem 6.2.4. *Let $(\Omega, H, \langle\cdot,\cdot\rangle, \|\cdot\|, U, K)$ be a strict SRKHS and $X = \{x_1, x_2, \ldots, x_n\} \subseteq \Omega$ be unisolvent w.r.t U. $\widehat{E}_X^{\boldsymbol{w},\boldsymbol{d}}$, defined as above, is continuous and strictly convex over $\mathcal{K}/U$.*

Proof. As $\mathcal{K}/U$ is a real finite dimensional normed space, to show continuity over $\mathcal{K}/U$, it suffices to show that $\widehat{E}_X^{\boldsymbol{w},\boldsymbol{d}}$ is convex over $\mathcal{K}/U$. the function $f+U \to (f+U,f+U)$ is strictly convex over $\mathcal{K}/U$. Let $\widehat{F}_X^{\boldsymbol{w},\boldsymbol{d}} : \mathcal{K}/U \to \mathbb{R}$ such that

$$\widehat{F}_X^{\boldsymbol{w},\boldsymbol{d}}(f+U) = \min_{g\in f+U}(\sum_{i=1}^n w_i(g(x_i)-d_i)^2)$$
$$= \min_{g\in f+U}[g-s,g-s],\ \forall f+U \in \mathcal{K}/U.$$

$f \to [f,f]$ is a convex function and as such $f \to [f-s,f-s]$ is the translation of a convex function and so is convex too. Let $f_1+U, f_2+U \in \mathcal{K}/U$, $\theta \in (0,1)$. Let $g_1 \in f_1+U, g_2 \in f_2+U$ be the unique minimisers of $E_X^{\boldsymbol{w},\boldsymbol{d}}$ in f_1+U, f_2+U respectively. So

$$\begin{aligned}\widehat{F}_X^{\boldsymbol{w},\boldsymbol{d}}((\theta f_1+(1-\theta)f_2)+U) &\le [\theta g_1+(1-\theta)g_2-s, \theta g_1+(1-\theta)g_2-s]\\ &\le \theta[g_1-s,g_1-s]+(1-\theta)[g_2-s,g_2-s]\\ &= \theta\widehat{F}_X^{\boldsymbol{w},\boldsymbol{d}}(f_1+U)+(1-\theta)\widehat{F}_X^{\boldsymbol{w},\boldsymbol{d}}(f_2+U).\end{aligned}$$

So $\widehat{F}_X^{\boldsymbol{w},\boldsymbol{d}}$ is convex. Hence $\widehat{E}_X^{\boldsymbol{w},\boldsymbol{d}}$ is the sum of a strictly convex and a convex function, making $\widehat{E}_X^{\boldsymbol{w},\boldsymbol{d}}$ a strictly convex function. □

Now let's prove the first of two main results in this section.

Theorem 6.2.5. *Let $(\Omega, H, \langle\cdot,\cdot\rangle, \|\cdot\|, U, K)$ be a strict SRKHS and let $E_X^{\boldsymbol{w},\boldsymbol{d}} : H \to \mathbb{R}$ be defined as follows.*

$$E_X^{\boldsymbol{w},\boldsymbol{d}}(f) = \langle f,f\rangle + \sum_{i=1}^n w_i(f(x_i)-d_i)^2,\ \ \forall f \in H,$$

where $\boldsymbol{w} = (w_1, w_2, \ldots, w_n) \in \mathbb{R}^n, \boldsymbol{w} > \boldsymbol{0}$, and $\boldsymbol{d} = (d_1, d_2, \ldots, d_n) \in \mathbb{R}^n$. There exists a unique minimiser of $E := E_X^{\boldsymbol{w},\boldsymbol{d}}$ over H.

Proof. Let $f \in H$. As our setting is a strict SRKHS, by Theorem 6.2.1, consider the unique minimal norm interpolant of f, $\mathcal{I}_X f \in \mathcal{K}$. $E(f) \ge E(\mathcal{I}_X f)$, with equality only when $f = \mathcal{I}_X f$, so if there exists a minimiser of E in $\mathcal{K}$, then said minimiser is a minimiser of E in H. Now let's consider the quotient space $\mathcal{K}/U$

with the natural energy $\widehat{E} := \widehat{E}_X^{\boldsymbol{w},\boldsymbol{d}}$. Let $s \in \mathcal{K}$ such that $s(x_i) = d_i, \ \forall i = 1, \dots, n$ and consider the set $V = \{f + U \in \mathcal{K}/U \mid \widehat{E}(f+U) \leq \widehat{E}(s+U) = (s+U, s+U)\}$. V is a subset of the set $\{f + U \in \mathcal{K}/U \mid (f+U, f+U) \leq (s+U, s+U)\}$ which is a closed, bounded set in a finite dimensional space $\mathcal{K}/U$, and so is compact. As $\widehat{E}$ is continuous, V is a closed subset in this compact set and as such is also compact. As $\widehat{E}$ is continuous, it achieves a minimum (say $s^* + U$) in V. This minimum is clearly a local minimum, as if it were not, then for any neighborhood of $s^* + U$, there would be a affine subset $g + U$ of lower energy, yet $g+U$ would be in V and as such $\widehat{E}(g+U) \geq \widehat{E}(s^*+U)$ which is a contradiction. The set V is also convex, to see this, first note that from Theorem 6.2.4, $\widehat{E}$ is strictly convex. Let $f_1 + U, f_2 + U \in V$. Let $\theta \in (0,1)$,

$$\begin{aligned}\widehat{E}(\theta(f_1+U) + (1-\theta)(f_2+U)) &\leq \theta \widehat{E}(f_1+U) + (1-\theta)\widehat{E}(f_2+U) \\ &\leq \theta \widehat{E}(s+U) + (1-\theta)\widehat{E}(s+U) = \widehat{E}(s+U).\end{aligned}$$

Thus $\theta(f_1+U)+(1-\theta)(f_2+U) \in V$. As $\widehat{E}$ is strictly convex over $\mathcal{K}/U$, the local minimum s^*+U of $\widehat{E}$ is a unique global minimum over $\mathcal{K}/U$. This implies that all minimisers of E in H, if any exist, must be in s^*+U. From Theorem 6.2.3, we have an element of the form $s^* + u, \ u \in U$, that uniquely minimises E in $s^* + U$. $s^* + u$ then minimises E over the whole of H. □

Now let's consider finding the unique minimiser of $E_X^{\boldsymbol{w},\boldsymbol{d}}$. The unique solution (say s^*) to this problem is of the form

$$s^* = \sum_{i=1}^{n} \lambda_i K_{x_i} + u^*$$

where $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \dots, \lambda) \in U_X^{\perp}, \ u^* \in U$. Let $\boldsymbol{W}$ is the $n \times n$ diagonal matrix with diagonal entries $(\boldsymbol{W})_{ii} = w_i \ \ \forall i = 1, \dots, n$. Let $B = \{u_1, u_2, \dots, u_m\}$ be a basis of U and $\boldsymbol{c} = (c_1, c_2, \dots, c_m) \in \mathbb{R}^m$ be the coefficient for u^* w.r.t B. Applying $E_X^{\boldsymbol{w},\boldsymbol{d}}$ to s^*, gives:

$$\begin{aligned}E_X^{\boldsymbol{w},\boldsymbol{d}}(s^*) &= \langle s^*, s^* \rangle + \sum_{i=1}^{n} w_i (s^*(x_i) - d_i)^2 \\ &= \langle \sum_{i=1}^{n} \lambda_i K_{x_i}, \sum_{j=1}^{n} \lambda_j K_{x_j} \rangle + \sum_{k=1}^{n} w_k (\sum_{l=1}^{n} \lambda_l K_{x_l}(x_k) + u^* - d_k)^2 \\ &= \sum_{i,j=1}^{n} \lambda_i \lambda_j K(x_i, x_j) + \sum_{k=1}^{n} w_k (\sum_{l=1}^{n} \lambda_l K(x_k, x_l) + \sum_{i=1}^{m} c_i u_i(x_k) - d_k)^2 \\ &= \boldsymbol{\lambda}^T \boldsymbol{K}_X \boldsymbol{\lambda} + (\boldsymbol{K}_X \boldsymbol{\lambda} + \boldsymbol{B}_X \boldsymbol{c} - \boldsymbol{d})^T \boldsymbol{W} (\boldsymbol{K}_X \boldsymbol{\lambda} + \boldsymbol{B}_X \boldsymbol{c} - \boldsymbol{d}) \\ &= \boldsymbol{\lambda}^T \boldsymbol{K}_X \boldsymbol{\lambda} + \|\boldsymbol{K}_X \boldsymbol{\lambda} + \boldsymbol{B}_X \boldsymbol{c} - \boldsymbol{d}\|_{\boldsymbol{W}}^2,\end{aligned}$$

where $\| \cdot \|_{\boldsymbol{W}} : \mathbb{R}^n \to \mathbb{R}$ is the norm $\|\boldsymbol{v}\|_{\boldsymbol{W}} = \sqrt{\boldsymbol{v}^t \boldsymbol{W} \boldsymbol{v}}, \ \forall \boldsymbol{v} \in \mathbb{R}^n$.

Define the two functions $\mathcal{E}^{\boldsymbol{w},\boldsymbol{d}}_{X,\boldsymbol{c}} : U^{\perp}_X \to \mathbb{R}$ and $\mathcal{E}^{\boldsymbol{w},\boldsymbol{d}}_{X,\boldsymbol{\lambda}} : \mathbb{R}^m \to \mathbb{R}$ as follows:

$$\mathcal{E}^{\boldsymbol{w},\boldsymbol{d}}_{X,\boldsymbol{c}}(\boldsymbol{a}) = \boldsymbol{a}^T \boldsymbol{K}_X \boldsymbol{a} + \|\boldsymbol{K}_X \boldsymbol{a} + \boldsymbol{B}_X \boldsymbol{c} - \boldsymbol{d}\|^2_{\boldsymbol{W}},\ \forall \boldsymbol{a} \in U^{\perp}_X,$$
$$\mathcal{E}^{\boldsymbol{w},\boldsymbol{d}}_{X,\boldsymbol{\lambda}}(\boldsymbol{c'}) = \boldsymbol{\lambda}^T \boldsymbol{K}_X \boldsymbol{\lambda} + \|\boldsymbol{K}_X \boldsymbol{\lambda} + \boldsymbol{B}_X \boldsymbol{c'} - \boldsymbol{d}\|^2_{\boldsymbol{W}},\ \forall \boldsymbol{c'} \in \mathbb{R}^m.$$

Given the equivalence of these functions to $E^{\boldsymbol{w},\boldsymbol{d}}_X$, they both have local minimisers ($\boldsymbol{\lambda}$ for $\mathcal{E}^{\boldsymbol{w},\boldsymbol{d}}_{X,\boldsymbol{c}}$ and $\boldsymbol{c}$ for $\mathcal{E}^{\boldsymbol{w},\boldsymbol{d}}_{X,\boldsymbol{\lambda}}$). As the gradients of these functions exists, the coefficient vectors $\boldsymbol{\lambda}$ and $\boldsymbol{c}$ to minimisation problem satisfies the first order nessesary conditions:

$$\nabla \mathcal{E}^{\boldsymbol{w},\boldsymbol{d}}_{X,\boldsymbol{c}}(\boldsymbol{\lambda}) = \boldsymbol{0},$$
$$\nabla \mathcal{E}^{\boldsymbol{w},\boldsymbol{d}}_{X,\boldsymbol{\lambda}}(\boldsymbol{c}) = \boldsymbol{0}.$$

From the second equation

$$\begin{aligned}\nabla \mathcal{E}^{\boldsymbol{w},\boldsymbol{d}}_{X,\boldsymbol{\lambda}}(\boldsymbol{c}) = 2\boldsymbol{B}_X^T \boldsymbol{W} \boldsymbol{B}_X \boldsymbol{c} + 2\boldsymbol{B}_X^T \boldsymbol{W}(\boldsymbol{K}_X \boldsymbol{\lambda} - \boldsymbol{d}) = \boldsymbol{0}.\\ \boldsymbol{B}_X^T \boldsymbol{W}(\boldsymbol{K}_X \boldsymbol{\lambda} + \boldsymbol{B}_X \boldsymbol{c} - \boldsymbol{d}) = \boldsymbol{0}.\end{aligned}$$

Recalling that $\boldsymbol{a} \in U^{\perp}_X$ if and only if $\boldsymbol{B}_X^T \boldsymbol{a} = \boldsymbol{0}$, the above equation shows us that $\boldsymbol{K}_X \boldsymbol{\lambda} + \boldsymbol{B}_X \boldsymbol{c} - \boldsymbol{d} \in U^{\perp}_X$. We have the additional constraint that $\boldsymbol{B}_X^T \boldsymbol{\lambda} = \boldsymbol{0}$.

$$\begin{aligned}\nabla \mathcal{E}^{\boldsymbol{w},\boldsymbol{d}}_X(\boldsymbol{\lambda}) = 2\boldsymbol{K}_X \boldsymbol{\lambda} + 2\boldsymbol{K}_X \boldsymbol{W} \boldsymbol{K}_X \boldsymbol{\lambda} + 2\boldsymbol{K}_X \boldsymbol{W}(\boldsymbol{B}_X \boldsymbol{c} - \boldsymbol{d}) = \boldsymbol{0}.\\ 2\boldsymbol{K}_X(\boldsymbol{\lambda} + \boldsymbol{W}(\boldsymbol{K}_X \boldsymbol{\lambda} + \boldsymbol{B}_X \boldsymbol{c} - \boldsymbol{d})) = \boldsymbol{0}.\end{aligned}$$

given that $\boldsymbol{\lambda}$ and $\boldsymbol{K}_X \boldsymbol{\lambda} + \boldsymbol{B}_X \boldsymbol{c} - \boldsymbol{d} \in U^{\perp}_X$, and the conditionally strict positive definiteness of K w.r.t $U^{\perp}_X$, the above equation gives us

$$\begin{aligned}\boldsymbol{\lambda} + \boldsymbol{W}(\boldsymbol{K}_X \boldsymbol{\lambda} + \boldsymbol{B}_X \boldsymbol{c} - \boldsymbol{d}) = \boldsymbol{0},\\ (\boldsymbol{I} + \boldsymbol{W}\boldsymbol{K}_X)\boldsymbol{\lambda} + \boldsymbol{W}(\boldsymbol{B}_X \boldsymbol{c} - \boldsymbol{d}) = \boldsymbol{0}.\\ \boldsymbol{W}((\boldsymbol{W}^{-1} + \boldsymbol{K}_X)\boldsymbol{\lambda} + \boldsymbol{B}_X \boldsymbol{c} - \boldsymbol{d}) = \boldsymbol{0}.\\ (\boldsymbol{W}^{-1} + \boldsymbol{K}_X)\boldsymbol{\lambda} + \boldsymbol{B}_X \boldsymbol{c} = \boldsymbol{d}.\end{aligned}$$

From this, the following block matrix system is obtained:

$$\begin{bmatrix} \boldsymbol{W}^{-1} + \boldsymbol{K}_X & \boldsymbol{B}_X \\ \boldsymbol{B}_X^T & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{\lambda} \\ \boldsymbol{c} \end{bmatrix} = \begin{bmatrix} \boldsymbol{d} \\ \boldsymbol{0} \end{bmatrix}$$

The $(n+m) \times (n+m)$ is clearly symmetric, but is also invertible. To see this, let $\boldsymbol{C}_X$ be an $n \times m$ matrix with columns that span the orthogonal complement of the column space of $\boldsymbol{B}_X$. So $\boldsymbol{B}_X^T \boldsymbol{C}_X = \boldsymbol{0}$ and there exists a vector (say $\boldsymbol{\gamma}$) such that $\boldsymbol{\lambda} = \boldsymbol{C}_X \boldsymbol{\gamma}$. Left multiplying the equation $(\boldsymbol{W}^{-1} + \boldsymbol{K}_X)\boldsymbol{\lambda} + \boldsymbol{B}_X \boldsymbol{c} = \boldsymbol{d}$ by $\boldsymbol{C}_X^T$ gives

$$\boldsymbol{C}_X^T(\boldsymbol{W}^{-1} + \boldsymbol{K}_X)\boldsymbol{\lambda} + \boldsymbol{C}_X^T \boldsymbol{B}_X \boldsymbol{c} = \boldsymbol{C}_X^T(\boldsymbol{W}^{-1} + \boldsymbol{K}_X)\boldsymbol{C}_X \boldsymbol{\gamma} = \boldsymbol{C}_X^T \boldsymbol{d} \tag{12}$$

the matrix $\boldsymbol{A} := \boldsymbol{C}_X^T(\boldsymbol{W}^{-1}+\boldsymbol{K}_X)\boldsymbol{C}_X$ is invertible. let $\boldsymbol{v} \in \mathbb{R}^m$, $\boldsymbol{v}^T\boldsymbol{C}_X^T(\boldsymbol{W}^{-1}+\boldsymbol{K}_X)\boldsymbol{C}_X\boldsymbol{v} = (\boldsymbol{C}_X\boldsymbol{v})^T\boldsymbol{W}^{-1}(\boldsymbol{C}_X\boldsymbol{v}) + (\boldsymbol{C}_X\boldsymbol{v})^T\boldsymbol{K}_X(\boldsymbol{C}_X\boldsymbol{v}) > 0$ As $\boldsymbol{C}_X$ is of full rank, $\boldsymbol{W}^{-1}$ is a diagonal matrix of only positive diagonal entries and $\boldsymbol{K}_X$ is conditionally positive definite over $U_X^{\perp}$. Hence $\boldsymbol{A}$ is positive definite and so is invertible. Rearranging (2) gives

$$\boldsymbol{C}_X^T((\boldsymbol{W}^{-1}+\boldsymbol{K}_X)\boldsymbol{\lambda} - \boldsymbol{d}) = \boldsymbol{0},$$

showing that $(\boldsymbol{W}^{-1}+\boldsymbol{K}_X)\boldsymbol{\lambda} - \boldsymbol{d}$ is in the column space of $\boldsymbol{B}_X$. As $\boldsymbol{B}_X$ is of full rank, $\boldsymbol{c}$ is uniquely characterised by the equation

$$\boldsymbol{B}_X\boldsymbol{c} = \boldsymbol{d} - (\boldsymbol{W}^{-1}+\boldsymbol{K}_X)\boldsymbol{\lambda}.$$

The main result of this section has been established.

Theorem 6.2.6. *Consider the following minimisation problem.*

Let $(\Omega, H, \langle\cdot,\cdot\rangle, \|\cdot\|, U, K)$ be a strict SRKHS. Let $X = \{x_1, x_2, \ldots, x_n\} \subseteq \Omega$, X unisolvent w.r.t U. The function $E_X^{\boldsymbol{w},\boldsymbol{d}} : H \to \mathbb{R}$ defined as

$$E_X^{\boldsymbol{w},\boldsymbol{d}}(f) = \langle f, f\rangle + \sum_{i=1}^{n} w_i(f(x_i) - d_i)^2, \quad \forall f \in H. \tag{13}$$

where $\boldsymbol{d} = (d_1, d_2, \ldots, d_n), \boldsymbol{w} = (w_1, w_2, \ldots, w_n) \in \mathbb{R}^n, \boldsymbol{w} > \boldsymbol{0}$, has a unique minimiser $s \in H$. This minimiser has the form

$$s(\,\cdot\,) = \sum_{i=1}^{n} \lambda_i K_{x_i}(\,\cdot\,) + u^*$$

where $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_n) \in U_X^{\perp}$ and $u^* \in U$. Letting $B = \{u_1, u_2, \ldots, u_m\}$ be an arbitrary basis for U and $\boldsymbol{c} = (c_1, c_2, \ldots, c_m) \in \mathbb{R}^m$ such that $u* = \sum_{i=1}^{m} c_i u_i$, we have the following linear systems:

$$\begin{bmatrix} \boldsymbol{W}^{-1}+\boldsymbol{K}_X & \boldsymbol{B}_X \\ \boldsymbol{B}_X^T & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{\lambda} \\ \boldsymbol{c} \end{bmatrix} = \begin{bmatrix} \boldsymbol{d} \\ \boldsymbol{0} \end{bmatrix}$$

where $\boldsymbol{W}$ is the $n \times n$ diagonal matrix with diagonal entries $(\boldsymbol{W})_{ii} = w_i \;\; \forall i = 1, \ldots, n$. Furthermore, this linear system has a unique solution for $(\boldsymbol{\lambda}, \boldsymbol{c})$.

7 The Gaussian Functions and Higher Precision Computation

7.1 Introduction

Recall the class of Guassian functions given by

$$e^{-\alpha\|x\|^2},\ \alpha > 0.$$

The corresponding strictly positive definite functions are of the form

$$\Phi(x, y) = e^{-\alpha\|x-y\|^2},\ \alpha > 0.$$

Let $X = \{x_1, x_2, \dots, x_n\} \subseteq \Omega$. We are interested in interpolants of the form

$$\sum_{i=1}^{n} \lambda_i \Phi(x_i, \cdot)$$

such that

$$\sum_{i=1}^{n} \lambda_i \Phi(x_i, x_k) = d_k.$$

Recall $\mathbf{\Phi}_X$, the Gramian matrix of Φ at X. One can find $\boldsymbol{\lambda}$ via solving the the matrix equation

$$\mathbf{\Phi}_X \boldsymbol{\lambda} = \boldsymbol{d}.$$

When it comes to computing $\boldsymbol{\lambda}$, the conditioning of $\mathbf{\Phi}_X$ is of concern. Lower values of α increase the conditioning at a super linear rate.

Appendices

A Semi Inner Product Spaces

A.1 Introduction

Let V be a real real vector space. The function $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$ is a (real) **semi inner product** over V if the following holds:

(1) $\langle \boldsymbol{v}, \boldsymbol{w} \rangle = \langle \boldsymbol{w}, \boldsymbol{v} \rangle, \ \forall \boldsymbol{v}, \boldsymbol{w} \in V$.

(2) $\langle s\boldsymbol{v} + t\boldsymbol{w}, \boldsymbol{u} \rangle = s\langle \boldsymbol{v}, \boldsymbol{u} \rangle + t\langle \boldsymbol{w}, \boldsymbol{u} \rangle, \ \forall \boldsymbol{v}, \boldsymbol{u}, \boldsymbol{w} \in V, \ \forall s, t \in \mathbb{R}$.

(3) $\langle \boldsymbol{v}, \boldsymbol{v} \rangle \geq 0, \ \forall \boldsymbol{v} \in V$.

The vector space V equipped with a semi inner product is called a **semi inner product space**. The difference between semi inner product spaces and inner product spaces is that in an inner product space, the inequality in (3) is strict for all non-zero vectors. It should be clear that all inner product spaces are semi inner product spaces. The function $\| \cdot \| : V \to \mathbb{R}$ defined by

$$\|\boldsymbol{v}\| = \langle \boldsymbol{v}, \boldsymbol{v} \rangle^{1/2}, \ \forall \boldsymbol{v} \in V.$$

satisfies the following properties:

(1) $\|\boldsymbol{v}\| \geq 0, \ \forall \boldsymbol{v} \in V$.

(2) $\|s\boldsymbol{v}\| = |s|\|\boldsymbol{v}\|, \ \forall s \in \mathbb{R}, \ \forall \boldsymbol{v} \in V$.

(3) $\|\boldsymbol{v} + \boldsymbol{w}\| \leq \|\boldsymbol{v}\| + \|\boldsymbol{w}\|, \ \forall \boldsymbol{v}, \boldsymbol{w} \in V$.

Any function that satisfies the properties above over a real vector space is a **semi norm** over said vector space. The first two properties are trivial to prove. Proving property 3 can first be achieved by first proving the Cauchy-Schwarz inequality for semi inner products.

Theorem A.1.1. *Let $(V, \langle \cdot, \cdot \rangle)$ be a semi inner product space. For any $\boldsymbol{v}, \boldsymbol{w} \in V$*

$$|\langle \boldsymbol{v}, \boldsymbol{w} \rangle|^2 \leq \langle \boldsymbol{v}, \boldsymbol{v} \rangle \langle \boldsymbol{w}, \boldsymbol{w} \rangle.$$

Or equivalently

$$|\langle \boldsymbol{v}, \boldsymbol{w} \rangle| \leq \|\boldsymbol{v}\|\|\boldsymbol{w}\|.$$

Proof. Let $\boldsymbol{v}, \boldsymbol{w} \in V$.

$$\begin{bmatrix} s & t \end{bmatrix} \begin{bmatrix} \langle \boldsymbol{v}, \boldsymbol{v} \rangle & \langle \boldsymbol{v}, \boldsymbol{w} \rangle \\ \langle \boldsymbol{w}, \boldsymbol{v} \rangle & \langle \boldsymbol{w}, \boldsymbol{w} \rangle \end{bmatrix} \begin{bmatrix} s \\ t \end{bmatrix} = \langle s\boldsymbol{v} + t\boldsymbol{w}, s\boldsymbol{v} + t\boldsymbol{w} \rangle \geq 0, \ \forall s, t \in \mathbb{R}.$$

So the above 2×2 matrix is non-negative definite and so its determinant is greater than or equal to zero giving

$$\langle \boldsymbol{v}, \boldsymbol{v} \rangle \langle \boldsymbol{w}, \boldsymbol{w} \rangle - |\langle \boldsymbol{v}, \boldsymbol{w} \rangle|^2 \geq 0.$$

Rearranging gives the desired inequality. □

Lemma A.1.2. *Let $(V, \langle \cdot, \cdot \rangle)$ be a semi inner product space.*

$$\|\boldsymbol{v} + \boldsymbol{w}\| \leq \|\boldsymbol{v}\| + \|\boldsymbol{w}\|, \ \forall \boldsymbol{v}, \boldsymbol{w} \in V.$$

Proof. Let $\boldsymbol{v}, \boldsymbol{w} \in V$.

$$\begin{aligned} \|\boldsymbol{v} + \boldsymbol{w}\|^2 = \|\boldsymbol{v}\|^2 + 2\langle \boldsymbol{v}, \boldsymbol{w} \rangle + \|\boldsymbol{w}\|^2 &\leq \|\boldsymbol{v}\|^2 + 2\|\boldsymbol{v}\|\|\boldsymbol{w}\| + \|\boldsymbol{w}\|^2 \\ &= (\|\boldsymbol{v}_1\| + \|\boldsymbol{v}_2\|)^2. \end{aligned}$$

Taking square roots gives the desired inequality. □

The kernel of the semi norm $\|\cdot\|$ is the set U of all vectors $\boldsymbol{u} \in V$ such that $\langle \boldsymbol{u}, \boldsymbol{u} \rangle = 0$. The set U will also be referred to as the **null space** of V.

Theorem A.1.3. *Let $(V, \langle \cdot, \cdot \rangle)$ be a semi inner product space with null space U. For any $\boldsymbol{u} \in U$, $\langle \boldsymbol{v}, \boldsymbol{u} \rangle = 0$ for any $\boldsymbol{v} \in V$. Furthermore, U is a vector subspace of V.*

Proof. $\boldsymbol{0} \in U$, so U is non-empty. Let $\boldsymbol{u} \in U$, $\boldsymbol{v} \in V$. By Cauchy-Schwarz

$$|\langle \boldsymbol{v}, \boldsymbol{u} \rangle|^2 \leq \langle \boldsymbol{v}, \boldsymbol{v} \rangle \langle \boldsymbol{u}, \boldsymbol{u} \rangle = \langle \boldsymbol{v}, \boldsymbol{v} \rangle \cdot 0 = 0.$$

So $\langle \boldsymbol{v}, \boldsymbol{u} \rangle = 0$. Now let $s, t \in \mathbb{R}$ and $\boldsymbol{u}' \in U$.

$$\langle \boldsymbol{v}, s\boldsymbol{u} + t\boldsymbol{u}' \rangle = s\langle \boldsymbol{v}, \boldsymbol{u} \rangle + t\langle \boldsymbol{v}, \boldsymbol{u}' \rangle = 0.$$

Hence $s\boldsymbol{u} + t\boldsymbol{u}' \in U$. □

It follows immediately that for any $\boldsymbol{v} \in V, \boldsymbol{v} \perp U$. From now onward in this section, $\boldsymbol{u}$ will be used to refer to elements of the null space U.

Lemma A.1.4. *Let $(V, \langle \cdot, \cdot \rangle)$ be a semi inner product space with null space U. The following hold*

(1) $\langle \boldsymbol{v}_1 + \boldsymbol{u}_1, \boldsymbol{v}_2 + \boldsymbol{u}_2 \rangle = \langle \boldsymbol{v}_1, \boldsymbol{v}_2 \rangle, \ \forall \boldsymbol{v}_1, \boldsymbol{v}_2 \in V, \ \forall \boldsymbol{u}_1, \boldsymbol{u}_2 \in U.$

(2) $\|\boldsymbol{v} + \boldsymbol{u}\| = \|\boldsymbol{v}\|, \ \forall \boldsymbol{v} \in V, \ \boldsymbol{u} \in U.$

Proof. Let $\boldsymbol{v}_1, \boldsymbol{v}_2 \in V$, $\boldsymbol{u}_1, \boldsymbol{u}_2 \in U$.

$$\langle \boldsymbol{v}_1 + \boldsymbol{u}_1, \boldsymbol{v}_2 + \boldsymbol{u}_2 \rangle = \langle \boldsymbol{v}_1, \boldsymbol{v}_2 \rangle + \langle \boldsymbol{v}_1, \boldsymbol{u}_2 \rangle + \langle \boldsymbol{u}_1, \boldsymbol{v}_2 \rangle + \langle \boldsymbol{u}_1, \boldsymbol{u}_2 \rangle = \langle \boldsymbol{v}_1, \boldsymbol{v}_2 \rangle.$$

Giving *1*.

$$\|\boldsymbol{v}_1 + \boldsymbol{u}_1\|^2 = \|\boldsymbol{v}_1\|^2 + \langle \boldsymbol{v}_1, \boldsymbol{u}_1 \rangle + \|\boldsymbol{u}_1\|^2 = \|\boldsymbol{v}_1\|^2$$

Taking square roots gives *2*. □

The topology induced by the semi norm $\|\cdot\|$ can be seen as degenerate as converging sequences in a semi norm space to not converge uniquely when the null space U is non-trivial ($U \neq \{\mathbf{0}\}$). If a given sequence $\{\boldsymbol{v}_i\}_{i\in\mathbb{N}}$ converges to $\boldsymbol{v}$ then as

$$\|\boldsymbol{v} - \boldsymbol{v}_i\| = \|(\boldsymbol{v} + \boldsymbol{u}) - \boldsymbol{v}_i\|, \ \boldsymbol{u} \in U,$$

the sequence $\{\boldsymbol{v}_i\}_{i\in\mathbb{N}}$ converges to every element in $\boldsymbol{v}+U = \{\boldsymbol{v}+\boldsymbol{u} \in V \mid \boldsymbol{u} \in U\}$. From the open ball perspective of the topology, every non-empty open ball $\mathcal{B}$ is such that $\mathcal{B} = \mathcal{B} + U = \{\boldsymbol{b} + \boldsymbol{u} \in V \mid \boldsymbol{b} \in \mathcal{B}, \boldsymbol{u} \in U.\}$.

The quotient space V/U (V mod U) is the set of all affine subsets of V in the form $\boldsymbol{v} + U$ for some $\boldsymbol{v} \in V$. Equivalently, V/U is a set of equivalence classes of the equivalence relation $\sim$ where $\boldsymbol{v}_1 \sim \boldsymbol{v}_2$ if and only if $\boldsymbol{v}_1 - \boldsymbol{v}_2 \in U$ for any $\boldsymbol{v}_1, \boldsymbol{v}_2 \in V$. This set is indeed a vector space with addition defined by $A + B = \{\boldsymbol{a} + \boldsymbol{b} \in V \mid \boldsymbol{a} \in A, \ \boldsymbol{b} \in B\}$ and scalar multiplication given by $sA = \{s\boldsymbol{a} \mid \boldsymbol{a} \in A\}$, for any $A, B \in V/U$ and $s \in \mathbb{R}$. U acts as the zero vector of V/U. The quotient space V/U has a natural inner product (say $(\cdot,\cdot)$) such that

$$(\boldsymbol{v}_1 + U, \boldsymbol{v}_2 + U) = \langle \boldsymbol{v}_1, \boldsymbol{v}_2 \rangle.$$

This inner product is well defined. To see this, let $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{w}_1, \boldsymbol{w}_2 \in V$ such that

$$\boldsymbol{v}_1 + U = \boldsymbol{w}_1 + U, \ \boldsymbol{v}_2 + U = \boldsymbol{w}_2 + U.$$

This implies that $\boldsymbol{w}_1 - \boldsymbol{v}_1, \ \boldsymbol{w}_2 - \boldsymbol{v}_2 \in U$.

$$\begin{aligned}
(\boldsymbol{v}_1 + U, \boldsymbol{v}_2 + U) &= \langle \boldsymbol{v}_1, \boldsymbol{v}_2 \rangle \\
&= \langle \boldsymbol{v}_1, \boldsymbol{v}_2 \rangle + \langle \boldsymbol{v}_1, \boldsymbol{w}_2 - \boldsymbol{v}_2 \rangle \\
&= \langle \boldsymbol{v}_1, \boldsymbol{w}_2 \rangle \\
&= \langle \boldsymbol{v}_1, \boldsymbol{w}_2 \rangle + \langle \boldsymbol{w}_1 - \boldsymbol{v}_1, \boldsymbol{w}_2 \rangle \\
&= \langle \boldsymbol{w}_1, \boldsymbol{w}_2 \rangle = (\boldsymbol{w}_1 + U, \boldsymbol{w}_2 + U).
\end{aligned}$$

It is also indeed an inner product as

$$(\boldsymbol{v} + U, \boldsymbol{v} + U) = \langle \boldsymbol{v}, \boldsymbol{v} \rangle = 0 \text{ if and only if } \boldsymbol{v} \in U,$$

implying that $\boldsymbol{v} + U = U$.

A.2 Semi Hilbert Spaces

What follows here on semi Hilbert spaces must be known but I have been unable to find a suitable reference. A given semi inner product space $(V, \langle\cdot,\cdot\rangle)$ with corresponding semi norm $\|\cdot\|$ and null space U is a **semi Hilbert space** if the quotient space V/U is a Hilbert space when equipped with its natural inner product $(\cdot,\cdot)$ as defined above. Every semi inner product space has a completion to a semi Hilbert space that is isomorphic to $\overline{V/U} \oplus U$ with the semi inner product $[\cdot,\cdot]$ given by

$$[(A, \boldsymbol{u}_1), (B, \boldsymbol{u}_2)] = (A, B), \ \forall A, B \in \overline{V/U}, \ \boldsymbol{u}_1, \boldsymbol{u}_2 \in U.$$

A projection P from V onto U is trivially an orthogonal projection as $P\boldsymbol{v} \in U$ and so $\langle P\boldsymbol{v}, (I-P)\boldsymbol{v}\rangle = 0$ from Theorem A.1.3, where I is the identity map over V. This means that for any projection P onto U, there is a corresponding orthogonal decomposition of V given by $V = (I-P)V \oplus PV = (I-P)V \oplus U$.

The set of all bounded linear functionals V^* has a rather degenerate property that for any $\mu \in V^*$, $\boldsymbol{u} \in U$, $|\mu\boldsymbol{u}| \leq c\|\boldsymbol{u}\| = 0$, so $\mu\boldsymbol{u} = 0$, that is to say, every bounded functional annihilates the null space U. The vector spaces of functionals V^* and $(V/U)^*$ can be seen as equivalent as for any $\mu \in V^*$ there exists a $\mu' \in (V/U)^*$ such that $\mu'(\boldsymbol{v}+U) := \mu\boldsymbol{v}$. This is because

$$|\mu'(\boldsymbol{v}+U)| = |\mu\boldsymbol{v}| \leq c\|\boldsymbol{v}\| = c|\boldsymbol{v}+U|, \text{ for some } c \in \mathbb{R},$$

where $|\boldsymbol{v}+U|$ is the norm of $\boldsymbol{v}+U$ with resepct to $(\cdot,\cdot)$. So $\mu' \in (V/U)^*$. A similar argument establishes that for any functional in $\mu' \in (V/U)^*$, there is a corresponding functional in $\mu \in V^*$ such that $\mu\boldsymbol{v} := \mu'(\boldsymbol{v}+U)$.

$$|\mu\boldsymbol{v}| = |\mu'(\boldsymbol{v}+U)| \leq c|\boldsymbol{v}+U| = c\|\boldsymbol{v}\|, \text{ for some } c \in \mathbb{R}.$$

Riesz representation theorem tell us that there exists a representation mapping $R : (V/U)^* \to V/U$. This can be used to induced a representation map for V by first taking an arbitrary projection P onto U. Consider the vector subspace $(I-P)V$ of V. There is a natural one-to-one correspondence between V/U and $(I-P)V$ as every element of $(I-P)V$ must exist in one element of V/U yet no two distinct elements $\boldsymbol{v}_1, \boldsymbol{v}_2 \in (I-P)V$ can be in the same set in V/U as that would imply that $\boldsymbol{v}_1 - \boldsymbol{v}_2 \in U$ yet this can only hold when $\boldsymbol{v}_1 - \boldsymbol{v}_2 = 0$. The mapping $R_P : V^* \to (I-P)V$ defined as follows

$$R_P(\mu) \in R(\mu') \cap (I-P)V,$$

where $\mu' \in (V/U)^*$ is the functional corresponding to $\mu \in V^*$. The set $R(\mu') \cap (I-P)V$ only has one element as established above. From the above we can reason that

$$\begin{aligned} R_P(\mu) + U &\subseteq R(\mu') \cap (I-P)V + U \\ &\subseteq (R(\mu') + U) \cap ((I-P)V + U) \\ &= R(\mu') \cap V = R(\mu'), \end{aligned}$$

As both $R_P(\mu) + U, R(\mu') \in V/U$ intersect (They both at least contain $R_P(\mu)$), and they are contained in a set of disjoint sets, they must be equal. Letting $\mu \in V^*$

$$\langle \boldsymbol{v}, R_P(\mu)\rangle = (\boldsymbol{v}+U, R_P(\mu)+U) = (\boldsymbol{v}+U, R(\mu')) = \mu'(\boldsymbol{v}+U) = \mu\boldsymbol{v}.$$

This established that R_P is a representation mapping for V.

Finally, the degeneracy of V^* was noted before, one can wonder if there could be a way to extend the bounded continuous functionals to include functionals that do not annihilate U. The following method can be found in [11]. Let V^{*P} be the set of functionals μ over V such that $\mu(I-P) \in V^*$. It should be clear that $V^* \subset V^{*P}$. R_P can be extended to V^{*P} by $R_P(\mu) = R_P(\mu(I-P))$, $\forall \mu \in V^{*P}$. We then obtain

$$\mu \boldsymbol{v} = \mu(I-P)\boldsymbol{v} + \mu P \boldsymbol{v} = \langle \boldsymbol{v}, R_P(\mu) \rangle + \mu P \boldsymbol{v}.$$

B Convex Sets and Convex Functions

Let V be a real vector space. The natural setting for the study of a convex function is a convex subset of V, so before defining convex functions, let's define convex sets. A subset S of V is **convex** if for any $\boldsymbol{v}, \boldsymbol{u} \in V$, the line connecting them is in S, that is to say, for any $\theta \in (0,1)$

$$\theta \boldsymbol{v} + (1-\theta)\boldsymbol{u} \in S.$$

Let $C : V \to \mathbb{R}$. C is **convex** over V (or just convex) if the following inequality holds:

$$C(\theta \boldsymbol{v} + (1-\theta)\boldsymbol{u}) \leq \theta C(\boldsymbol{v}) + (1-\theta) C(\boldsymbol{u})$$

where $\boldsymbol{v}, \boldsymbol{u} \in V$, $\theta \in (0,1)$. If C is convex over V then C is called **strictly convex** over V (or just strictly convex) if the above inequality is strict when $v \neq u$. Only the theorems essential for this thesis are included here. For a reference, see [10].

Theorem B.0.1. *Let V be a real vector space. Let C be a (strictly) convex function over a real vector space V. Any translation of C (i.e. $C_{\boldsymbol{c}} : V \to \mathbb{R}$ such that $C_{\boldsymbol{c}}(\cdot) = C(\cdot + \boldsymbol{c})$ for some $\boldsymbol{c} \in V$) is (strictly) convex over V.*

Proof. Let $\boldsymbol{c}$ and $C_{\boldsymbol{c}}$ be defined as above and let $\boldsymbol{v}, \boldsymbol{u} \in V$, $\boldsymbol{v} \neq \boldsymbol{u}$, $\theta \in (0,1)$.

$$\begin{aligned} C_{\boldsymbol{c}}(\theta \boldsymbol{v} + (1-\theta)\boldsymbol{u}) &= C(\theta \boldsymbol{v} + (1-\theta)\boldsymbol{u} + \boldsymbol{c}) \\ &= C(\theta(\boldsymbol{v} + \boldsymbol{c}) + (1-\theta)(\boldsymbol{u} + \boldsymbol{c})) \\ &\leq \theta C(\boldsymbol{v} + \boldsymbol{c})) + (1-\theta) C(\boldsymbol{u} + \boldsymbol{c}) \\ &= \theta C_{\boldsymbol{c}}(\boldsymbol{v}) + (1-\theta) C_{\boldsymbol{c}}(\boldsymbol{u}). \end{aligned}$$

So $C_{\boldsymbol{c}}$ is convex. If C is strictly convex then the above inequality is strict and so $C_{\boldsymbol{c}}$ is strictly convex. $\square$

Theorem B.0.2. *Let V be a real vector space. Let C_1 and C_2 be convex functions over V. $C_1 + C_2$ is convex. Furthermore, if either C_1 and/or C_2 is strictly convex, then $C_1 + C_2$ is strictly convex.*

Proof. Let $\boldsymbol{v}, \boldsymbol{u} \in V$, $\boldsymbol{v} \neq \boldsymbol{u}$, $\theta \in (0,1)$.

$$\begin{aligned} (C_1 + C_2)(\theta \boldsymbol{v} + (1-\theta)\boldsymbol{u}) &\leq \theta C_1(\boldsymbol{v}) + (1-\theta) C_1(\boldsymbol{u}) + \theta C_2(\boldsymbol{v}) + (1-\theta) C_2(\boldsymbol{u}) \\ &= \theta (C_1 + C_2)(\boldsymbol{v}) + (1-\theta)(C_1 + C_2)(\boldsymbol{u}). \end{aligned}$$

If either f or g is strictly convex, then the above inequality is strict. $\square$

Theorem B.0.3. *Let $(V, \langle \cdot, \cdot \rangle)$ be a real semi inner product space. The function $C : V \to \mathbb{R}$ given by $C(\boldsymbol{v}) = \langle \boldsymbol{v}, \boldsymbol{v} \rangle$ for all $v \in V$ is convex. Furthermore, if $(V, \langle \cdot, \cdot \rangle)$ is an inner product space, C is strictly convex.*

Proof. Denote by $\|\cdot\|$ the induced norm from the inner product. $C(\cdot) = \|\cdot\|^2$. Let $\boldsymbol{v}, \boldsymbol{u} \in V,\ \boldsymbol{v} \neq \boldsymbol{u},\ \theta \in (0,1)$. Proving the convexity of C is equivalent to proving the following

$$\theta C(\boldsymbol{v}) + (1-\theta)C(\boldsymbol{u}) - C(\theta\boldsymbol{v} + (1-\theta)\boldsymbol{u}) \geq 0.$$

Given this

$$\begin{aligned}
&\theta C(\boldsymbol{v}) + (1-\theta)C(\boldsymbol{u}) - C(\theta\boldsymbol{v} + (1-\theta)\boldsymbol{u}) \\
&= \theta\|\boldsymbol{v}\|^2 + (1-\theta)\|\boldsymbol{u}\|^2 - (\theta^2\|\boldsymbol{v}\|^2 + (1-\theta)^2\|\boldsymbol{u}\|^2 + 2\theta(1-\theta)\langle\boldsymbol{v},\boldsymbol{u}\rangle) \\
&= \theta(1-\theta)\|\boldsymbol{v}\|^2 + \theta(1-\theta)\|\boldsymbol{u}\|^2 - 2\theta(1-\theta)\langle\boldsymbol{v},\boldsymbol{u}\rangle \\
&= \theta(1-\theta)(\|\boldsymbol{v}\|^2 + \|\boldsymbol{u}\|^2 - 2\langle\boldsymbol{v},\boldsymbol{u}\rangle) \\
&\geq \theta(1-\theta)(\|\boldsymbol{v}\| - \|\boldsymbol{u}\|)^2 \geq 0.
\end{aligned}$$

So $\theta C(\boldsymbol{v}) + (1-\theta)C(\boldsymbol{u}) \geq C(\theta\boldsymbol{v} + (1-\theta)\boldsymbol{u})$, that is to say, C is convex. There are two inequalities above, if either of them are strict, it would follow that C is strictly convex. It will be shown that in the case where $\langle\cdot,\cdot\rangle$ is a strict inner product (which we now suppose), at least one of the above inequalities are strict. By Cauchy-Schwarz the equation

$$\|\boldsymbol{v}\|^2 + \|\boldsymbol{u}\|^2 - 2\langle\boldsymbol{v},\boldsymbol{u}\rangle = (\|\boldsymbol{v}\| - \|\boldsymbol{u}\|)^2$$

holds only when $\boldsymbol{v}$ and $\boldsymbol{u}$ are linearly dependent. When $\boldsymbol{v}$ and $\boldsymbol{u}$ are linearly dependent, then is trivially follows that

$$(\|\boldsymbol{v}\| - \|\boldsymbol{u}\|)^2 > 0,$$

given that $\boldsymbol{v}$ and $\boldsymbol{u}$ are distinct. So in all cases $\theta C(\boldsymbol{v}) + (1-\theta)C(\boldsymbol{u}) > C(\theta\boldsymbol{v} + (1-\theta)\boldsymbol{u})$. Hence C is strictly convex. □

The following theorem is from Cheney (1996) [5].

Theorem B.0.4. *Let V be a real finite dimensional normed space. Let C be a convex function over V. C is continuous over V.*

Lemma B.0.5. *Let V be a real vector space. Let C be a convex function over over a convex subset $S \subseteq V$. If $\boldsymbol{v}, \boldsymbol{u} \in V$ are local minimisers of C in S, $C(u) = C(v)$.*

Proof. Let $\boldsymbol{v}, \boldsymbol{u} \in V$ be local minimisers of C in S. Suppose that $C(\boldsymbol{u}) < C(\boldsymbol{v})$ and let $\theta \in (0,1)$. It follows that

$$C(\theta\boldsymbol{v} + (1-\theta)\boldsymbol{u}) \leq \theta C(\boldsymbol{v}) + (1-\theta)C(\boldsymbol{u}) < \theta C(\boldsymbol{u}) + (1-\theta)C(\boldsymbol{u}) = C(\boldsymbol{u}).$$

As S is convex, $\theta\boldsymbol{v} + (1-\theta)\boldsymbol{u} \in V$ for any $\theta \in (0,1)$. By letting θ approach zero the above strict inequality shows that $\boldsymbol{u}$ can not be a local minimiser. Contradiction, so $C(\boldsymbol{u}) \geq C(\boldsymbol{v})$. A similar argument establishes that $C(\boldsymbol{v}) \geq C(\boldsymbol{u})$. Hence $C(\boldsymbol{u}) = C(\boldsymbol{v})$. □

Theorem B.0.6. *Let V be a real vector space. Let C be a strictly convex function over a convex subset $S \subseteq V$. If there exists a local minimiser for C in S, then said local minimiser of C in S is the unique global minimiser over of C over S.*

Proof. Let $\boldsymbol{v}$ and $\boldsymbol{u}$ be local minimisers of C. Suppose $\boldsymbol{v} \neq \boldsymbol{u}$. Let $\theta \in (0,1)$, from Lemma B.0.5

$$C(\theta\boldsymbol{v} + (1-\theta)\boldsymbol{u}) < \theta C(\boldsymbol{v}) + (1-\theta)C(\boldsymbol{u}) = C(\boldsymbol{v}) = C(\boldsymbol{u}).$$

As S is convex, $\theta\boldsymbol{v} + (1-\theta)\boldsymbol{u} \in V$ for any $\theta \in (0,1)$. By letting θ approach either one or zero the above strict inequality shows that neither $\boldsymbol{v}$ or $\boldsymbol{u}$ can be local minimisers in S. Contradiction. Hence $\boldsymbol{v} = \boldsymbol{u}$. $\square$

References

[1] Aronszajn, N., Theory of Reproducing Kernels. *Transactions of the American Methematical Society*, 68:337-404, 1950.

[2] Bezhaev, A.Y., & Valisenko, V.A., Variational Theory of Splines, Springer, New York, 2001.

[3] Bingham, N.H., Positive definite functions on spheres. Proc. Camb. Phil. Soc. 73 (1973), 145-156.

[4] Boor, C.de., A Practical Guide to Splines, Springer, New York, 1978.

[5] Cheney, W., Introduction to Approximation Theory, McGraw-Hill, 1966.

[6] Cheney, W. & Light, W., A Course in Approximation Theory, Brooks/Cole, Pacific Grove, 2000.

[7] Davis, P.J., Interpolation and Approximation, Dover, New York, 1975.

[8] Duchon, J. Splines minimizing rotation-invariant semi-norms in Sobolev spaces, W. Schempp and K. Zeller Constructive (eds.), Theory of Functions of Several Variables, Lect. Notes Math. 571, Springer, Berlin, 1977 85-100.

[9] Mosamam, A.M., & Kent, J.T., Semi-reproducing kernel Hilbert spaces, splines and increment kriging, J. Nonparam. Stat. 22(6) (2010), 711-722.

[10] Niculescu, C.P., Persson, L.-E., Convex Functions and their Applications, Springer, New York, 2006.

[11] Schaback, R., Native spaces for radial basis functions I, in: M.W Müller, M.D. Buhmann, D.H. Mache, M. Felten (Eds.), New Developments in Approximation Theory, International Series of Numerical Mathematics, Vol. 132, Birkhäuser, Basel, 1999, pp. 255-282.

[12] Schoenberg, I.J. (1938), Metric spaces and completely monotone functions. Ann. Math. 39, 811-841.

[13] Schoenberg, I.J. (1942), Positive definite functions on spheres. Duke Math. J. 9, 96-108.

[14] Schumaker, L.L., Spline Functions: Basic Theory, Springer, Cambridge University Press, Cambridge, 2007.

[15] Strauß, H., Extremal properties of conditionally positive definite functions, C.K. Chui et al (eds.) Approximation Theory X: Wavelets, Splines and Applications, Vanderbilt University Press 2002, pp/ 441-452.

[16] Schölkopf, B, & Smola, A.J., Learning with Kernels, The MIT Press, Cambridge, 2002.

[17] Szegő, G., Orthogonal Polynomials. Amer. Math. Soc. Colloquium Publ. vol. 23, New York, 1959.

[18] Wendland, H., Scattered Data Approximation, Cambridge University Press, Cambridge, 2005.